HOPPING GREEN SAMS & SMITH
PROFESSIONAL ASSOCIATION
ATTORNEYS AND COUNSELORS
123 SOUTH CALHOUN STREET
POST OFFICE BOX 6526
TALLAHASSEE, FLORIDA 32314
(850) 222-7500
FAX (850) 224-8551
FAX (850) 425-3415
http://www.hgss.com

JAMES S. ALVES
BRIAN H. BIBEAU
RICHARD S. BRIGHTMAN
KEVIN B. COVINGTON
PETER C. CUNNINGHAM
RALPH A. DEMEO
WILLIAM H. GREEN
WADE L. HOPPING
GARY K. HUNTER, JR.
JONATHAN T. JOHNSON
LEIGH H. KELLETT
ROBERT A. MANNING
FRANK E. MATTHEWS
RICHARD D. MELSON
ANGELA R. MORRISON
SHANNON L. NOVEY
ERIC T. OLSEN

GARY V. PERKO
MICHAEL P. PETROVICH
DAVID L. POWELL
WILLIAM D. PRESTON
CAROLYN S. RAEPPLE
DOUGLAS S. ROBERTS
D. KENT SAFRIET
GARY P. SAMS
TIMOTHY G. SCHOENWALDER
ROBERT P. SMITH
DAN R. STENGLE
CHERYL G. STUART
W. STEVE SYKES

OF COUNSEL
ELIZABETH C. BOWMAN

ORIGINAL

Writer's Direct Dial No.
(850) 425-2313

November 22, 2000

RECEIVED-FPSC
00 NOV 22 PM 3:41
RECORDS AND REPORTING

BY HAND DELIVERY

Blanca Bayó
Director, Records and Reporting
Florida Public Service Commission
2540 Shumard Oak Boulevard
Tallahassee, FL 32399

Re: Docket No. 000121-TP

Dear Ms. Bayó:

Enclosed for filing on behalf of WorldCom, Inc. are the original and fifteen copies of its Comments Concerning Staff's Draft Performance Assessment Plan.

By copy of this letter, this document has been furnished to the parties on the attached service list.

If you have any questions regarding this filing, please call.

Very truly yours,

Richard D. Melson

RDM/mee
Enclosures
cc: Parties of Record

APP
CAF
CMP 3
COM 5
CTR
ECR 1
LEG 1
OPC
PAI
RGO Harvey
SEC 1
SER
OTH Harvey Done 11/30/00

RECEIVED & FILED
FPSC-BUREAU OF RECORDS

DOCUMENT NUMBER-DATE
15126 NOV 22 8
FPSC-RECORDS/REPORTING

BEFORE THE FLORIDA PUBLIC SERVICE COMMISSION

ORIGINAL

In re: Investigation into the Establishment of Operations Support Systems Permanent Performance Measures for Incumbent Local Exchange Telecommunications Companies	))))))	Docket No. 000121-TP Filed: November 22, 2000

WORLDCOM'S COMMENTS CONCERNING STAFF'S DRAFT PERFORMANCE ASSESSMENT PLAN

WorldCom, Inc. ("WorldCom") submits these comments in response to Staff's Draft Performance Assessment Plan ("Draft Plan").

INTRODUCTION

The Draft Plan provides a good starting point in the development of a Florida remedies plan. WorldCom supports a number of features in the plan, such as the requirement that BellSouth pay ALECs directly when it fails to meet performance standards; the implementation of a procedural, rather than an absolute cap on remedies; and the requirement that the Plan commence within a set time independently of whether BellSouth has been granted in-region long distance authority. There are a number of other parts of the Draft Plan, however, that WorldCom submits should be modified to ensure BellSouth is given the proper incentives to open its Florida local market. In that regard, WorldCom generally agrees with the comments being filed by AT&T on the Draft Plan. WorldCom files these separate comments to emphasize certain statistical and structural modifications that would improve the plan.[1]

[1] These comments are by no means exhaustive, and the focus on certain statistical and structual issues here is not intended to suggest that other issues are not also critical. Several other issues will be addressed in the

DOCUMENT NUMBER-DATE
15126 NOV 22 8
FPSC-RECORDS/REPORTING

COMMENTS

A. Statistical Modifications

WorldCom is most concerned about the choice of a parameter delta in the Draft Plan. WorldCom only has supported AT&T's parameter delta of .25 as the upper limit on what would be acceptable as a crude decision rule on competitive significance. For aggregate (i.e., Tier II) ALEC results, a .10 parameter delta should be chosen because of the larger sample sizes that would be involved. As the attached paper from Auburn University Economics Professor John D. Jackson notes, for larger sample sizes a large parameter delta can cut off major differences in means from the remedy scheme. As a result, the high parameter delta proposed in the Draft Plan would substantially limit the remedies BellSouth would be called upon to pay. Thus, a lower parameter delta should be adopted. Alternatively, competitively significant margins should be defined for each metric based actual market experience. These margins would have to be reexamined as competition develops and customers' reactions to differences in performance change.

B. Structural Modifications

WorldCom's strong preference is that remedies be assessed on a per measure basis rather than on a per occurrence basis. Per occurrence plans may work when competition is robust and few new products are coming to market, but in Florida, where competition is still struggling for a foothold, a per occurrence plan could generate low remedies that BellSouth readily would pay rather than open the doors to local competition. Per occurrence plans keep remedies the lowest when ALECs are just

---

testimony WorldCom will file in this docket. WorldCom also will address the issues raised in these comments more comprehensively in that testimony.

beginning to ramp up in a market or launching new services in competition with ILECs. ALEC reputations and financial resources are most vulnerable in those early stages of market entry or of a product offering. Competitors could be driven out of the market long before per occurrence remedies would reach levels to motivate BellSouth to spend money for human and capital resources to fix problems, let alone offset BellSouth's powerful incentive to retain existing local profits, new high-margin advanced digital service profits, and eventually long distance profits.

If the Commission nonetheless determines that a per occurrence plan should be implemented, a number of steps can be taken to ameliorate the problem of insufficient remedies during the early stages of competition. First, the Commission should require a minimum payment for each measure for which BellSouth fails to provide satisfactory performance. Such minimum payments would help give sufficient incentive for BellSouth to comply with its duty to provide parity and a meaningful opportunity to compete even when activity levels are low. Second, the Commission should increase the per occurrence remedies proposed in the Draft Plan. The base remedy amounts proposed are too low to provide an adequate incentive for BellSouth to cooperate with its competitors in the local market, and would have little impact on a company the size of BellSouth. Third, remedies should increase substantially for severe and repeated violations. The Draft Plan does not take into account the magnitude of poor performance by BellSouth, but rather only the number of customers that have been harmed. For example, the Draft Plan does not distinguish whether a performance standard was exceeded by 1 day for 100 customers or 30 days for 100 customers. In both instances the same remedy would apply. And although the Tier I remedy amounts do increase for repeated violations, those increases are not substantial enough to provide a sufficient incentive to provide good performance.

Structural problems also exist for Tier II. For example, Tier II remedy payments are not triggered unless BellSouth has discriminated against the entire ALEC community for three consecutive months. But even one month of poor performance, such as during an ALEC's ramp-up before it has established a reputation in the local market, can erode prospects for local competition. And it is difficult to imagine that two consecutive months of poor performance would not have a serious impact on an ALEC at any stage of market entry. Under the Draft Plan, it is possible for BellSouth to provide discriminatory service in eight out of twelve months and still pay no penalty. Thus, the Tier II remedies may rarely, if ever, be triggered, leaving BellSouth with only the prospect of paying Tier I remedies. Moreover, under Tier II (as under Tier I) BellSouth pays the same remedy regardless of the severity of the violation.

Finally, the Tier III remedy in the Draft Plan is too easy for BellSouth to avoid. So long as it did not fail any twelve or more of twenty-six performance standards for three consecutive months, BellSouth would remain free to market and sell long distance services, assuming it previously had been granted 271 authority. Thus, even if BellSouth provided atrocious performance on eleven performance measurements that thwarted ALECs' efforts to compete in the local market, the Tier III remedy would not be triggered. WorldCom respectfully submits that a more stringent test should be applied.

CONCLUSION

For the foregoing reasons, and those expressed in the Comments of AT&T, WorldCom respectfully requests the Commission to modify the Draft Plan. WorldCom will describe in more detail the modifications it believes are appropriate in the testimony it plans to file in this docket.

Respectfully submitted, this 22nd day of November, 2000.

Richard D. Melson
Hopping Green Sams & Smith, P.A.
P.O. Box 6526
Tallahassee, FL 32314
(850) 425-2313

Donna Canzano McNulty
WorldCom, Inc.
325 John Knox Road, Ste. 105
Tallahassee, FL 32303
(850) 422-1254

Dulaney L. O'Roark III
WorldCom, Inc.
Six Concourse Parkway
Suite 3200
Atlanta, GA 30328
(770) 284-5498

Attorneys for WorldCom, Inc.

CERTIFICATE OF SERVICE

I HEREBY CERTIFY that a copy of the foregoing was furnished to the following parties by U.S. Mail or Hand Delivery (*) this 22nd day of November, 2000.

Marsha Rule
AT&T Communications
101 North Monroe St. Suite 700
Tallahassee, FL 32301

Nancy B. White
c/o Nancy H. Sims
BellSouth Telecommunications, Inc.
150 S. Monroe St., Suite 400
Tallahassee, Fl 32301

Michael A. Gross
Florida Cable Telecommunications Assn.
246 E. 6th Avenue
Tallahassee, FL 32303

Kimberly Caswell
Verizon Select Services Inc.
P.O. Box 110, FLTC0007
Tampa, FL 33601

Scott Sapperstein
Intermedia Communications, Inc.
1 Intermedia Way
Tampa, FL 33647

Nanette Edwards/Brian Musselwhite
ITC^ Deltacom
4092 S. Memorial Parkway
Huntsville, AL 35802

Peter Dunbar/Karen Camechis
Pennington Law Firm
P.O. Box 10095
Tallahassee, FL 32302-2095

Mark Buechele
Supra Telecom
1311 Executive Center Drive, Suite 200
Tallahassee, FL 32301

Carolyn Marek
Time Warner Telecom of Florida, L.P.
233 Bramerton Court
Franklin, TN 37069

Tim Vaccaro
Division of Legal Services
Public Service Commission
2540 Shumard Oak Blvd.
Tallahassee, FL 32399

Susan Masterton/Charles Rehwinkel
Sprint Communications Company
P.O. Box 2214
MC: FLTLH00107
Tallahassee, FL 32316-2214

Catherine F. Boone
Covad Communications Company
10 Glenlake Parkway
Suite 650
Atlanta, Ga 30328-3495

Donna C. McNulty
MCI WorldCom
325 John Knox Road
The Atrium Suite 105
Tallahassee, Fl 32303

Patrick Wiggins
Charles Pellegrini
Katz Kutter Law Firm
106 E. College Ave. - 12th Floor
Tallahassee, FL 32301

Jeffrey Wahlen
Auslet Law Firm
P.O. Box 391
Tallahassee, FL 32302

Monica Barone
Birch Telecom of the South
8001 Fairlake Drive
Wake Forest, NC 27587

Jonathan Canis/Michael Hazzard
Kelley Law Firm
1200 19th St. NW, Fifth Floor
Washington, DC 20036

Joseph McGlothlin/Vicki Kaufman
McWhirter Law Firm
117 S. Gadsden Street
Tallahassee, FL 32301

Floyd Self
Messer Law Firm
P.O. Box 1876
Tallahassee, FL 32302

Kenneth Hoffman/John Ellis
Rutledge Law Firm
P.O. Box 551
Tallahassee, FL 32302

John Rubino/George S. Ford
Z-Tel Communications, Inc.
601 S. Harbour Island Blvd.
Tampa, FL 33602

Jeremy Marcus/Elizabeth Braman
Blumenfeld & Cohen
1625 Massachusetts Ave. NW, Suite 300
Washington, D.C. 20036

____
Attorney

139639.2

CRUCIAL SHORTCOMINGS OF THE "BALANCING CRITICAL VALUE" APPROACH TO PERFORMANCE APPRAISAL

by
John D Jackson, Professor of Economics,
Auburn University, Auburn, AL 36830

I. Introduction

Section 271 of the Telecommunication Act of 1996 provided for ILEC entry into the long distance telephone service market if CLECs were allowed to enter the various local telephone service markets. This CLEC entry, in turn, is predicated upon the CLECs' ability to purchase from the ILEC various services crucial to their ability to compete in the local market. Consequently, the Act further requires that the ILEC provide these services to the CLECs at a quality level *at least equal* to that they provide to their own customers or affiliates. Thus, the evaluation of parity in local service provision has become a central issue in all proceedings concerning ILECs' 271 approval. Statistical means difference tests, typically based on (some version of) the LCUG Z statistic, have become the cornerstone in the evaluation of service quality provision. Indeed, test results are not only used to determine whether the ILEC has discriminated against the CLEC in service quality provision, they also enter into the determination of the magnitude of the penalty involved according to several performance assurance plans (such as those proposed by SBT, BST, and AT&T). It is this latter use that has led to the development of a "balancing critical values" approach to parity testing and performance appraisal.

When one makes a decision concerning the presence or absence of parity in service provision based on a statistical test, he or she can err in one of two possible ways. They could conclude that discrimination in service provision exists when in fact it does not, or they could conclude that discrimination does not exist when in fact it does. Because the null hypothesis of the test assumes "no discrimination," the former error involves the rejection of a true null. It is called a type I error, and the probability (or risk) of committing such an error is called α. The latter error involves the acceptance of a false null. It is called a type II error, and the probability (or risk) of committing such an error is called β. The BCV approach to parity testing amounts to determining a critical value of the test statistic called a balancing critical value (BCV), that equates α with β. This principle was first enunciated by LCUG in the early (pre 1998) stages of parity testing discussions, but the current version is the result of joint efforts of BST's statistical discussions from Ernst and Young and AT&T's (now retired) statistical expert Colin Mallows. Indeed, a BCV has become an integral part of both AT&T and BST's Performance Assurance Plans (PAPs).

In principle, an equal chance of error approach is attractive for (at least) two reasons. First, it remedies a number of difficulties encountered by the alternative approach. A number of PAPs, e.g., SBT's Texas plan, employ a fixed critical value of the test statistic and a K-table in lieu of BCV. Without going into a detailed criticism, the K-table corrects for random variation in the test statistic by allowing the ILEC to fail "k" tests per month without penalty. Many CLECs object to this approach because the table is derived based on an unrealistic alternative (that the ILEC always provides parity service) and because it ignores type II errors. The BCV approach avoids these criticisms (and handles the random variation problem) by employing a critical value of the test statistic that equates the probabilities of committing type I and type II errors.

Second, the BCV approach dovetails neatly with the objective of unbiased penalty assessment. An optimal statistical decision would be one that equates the costs of making a type I error with the costs of making a type II error. ILEC representatives are typically more than willing to disclose how much a type I error costs them. CLECs, on the other hand, have a more difficult time determining how much a type II error costs them. These costs involve not only the foregone penalty payment and the cost to their reputation; they also entail the cost to society of having to continue monopolistic service provision while losing the benefits of competition. Since these costs are difficult to calculate, it is not reasonable to expect an optimal statistical decision. The BCV, however, accomplishes the next best thing. Since, the probability that the ILEC would have to pay a fine when it is not discriminating is equal to the probability that it will not have to pay a fine when it is discriminating, the long run expected value of inappropriate net penalty payments is zero.

It is indisputable that the BCV approach has a definite allure for parity testing and performance appraisal. Unfortunately, operationalizing the BCV approach, putting the principle into practice, exposes a major flaw which can open Pandora's Box in terms of allowing the ILEC to thwart meaningful CLEC competition at the local level. The problem relates to the key role played by a parameter δ in determining what critical values of the test statistic will lead to the rejection of parity. The flaw is that the value given to δ is arbitrarily determined; Pandora's Box is opened when δ is set equal to "large" values; and all the evidence suggests that ILECs are intent on pursuing exactly this strategy.

II. The Importance of Specifying Delta

To apply the BCV approach, one must (a) determine an expression for the value of α assuming the null hypothesis is true, (b) determine an expression for the value of β assuming the alternative hypothesis is true, and (c) set these two expressions equal to each other so as to solve for the balancing critical value (BCV) of the test statistic that equates α and β. Step (a) is easy because the CLEC and ILEC population means are assumed to be equal -- it does not matter what value they are equal to, just that they are equal to each other. The procedure becomes problematic at step (b) because we must have a specific value for the difference between the CLEC and ILEC population means in order to compute β. This is the point in the argument at which statisticians typically cop out. Ideally, we would like to compute β based on a means difference that is only just large enough to be marginally "competitively significant." Statisticians argue that they are in no position to gauge how large means differences should be in order to be marginally competitive significant, this matter should be left to "telephony experts." But given a measure of this difference, they can easily compute the BCV and hence implement an equal probability of Type I and Type II errors. The AT&T/BST statisticians capsulize the problem as follows:

$$H_0: \mu_C = \mu_I; \sigma_C^2 = \sigma_I^2$$
$$H_A: \mu_C = \mu_I + \delta \bullet \sigma_I; \sigma_I^2 = \lambda \sigma_C^2 \quad (1)$$

(Clearly, parity service provision requires both equality of means and equality of variances. The second set of equalities in H_0 and H_A above allow for discrimination in the form of the CLEC variance exceeding the ILEC variance by a multiplicative factor λ, $\lambda>1$; i.e., the ILEC provides the CLEC more variable service than it provides itself. While this is certainly an important source of discrimination, it is of only tangential importance to the problem at hand. Thus, in what follows, the variances are assumed to be equal; i.e., $\lambda=1$.) In this view, the CLEC and ILEC means are equal under H_0 and differ by an amount equal to $\delta \bullet \sigma_I$ under H_A. Analytically, under these assumptions, steps (a), (b), and (c) lead to the formula

$$BCV = \frac{\text{Expected Value of Test Statistic Assuming Means Differ by } \delta \cdot \sigma_I - \text{Expected Value of Test Statistic Assuming Means do not Differ}}{\text{Variance of Test Statistic Assuming Means Differ by } \delta \cdot \sigma_I + \text{Expected Value of Test Statistic Assuming Means do not Differ}} \quad (2)$$

Thus δ is a measure, in units of the ILEC standard deviation, of the extent to which the ILEC mean exceeds the CLEC mean (or, conversely). As such, specifying δ specifies the difference between the CLEC and ILEC means that would be marginally competitively significant in affecting local service competition. Further, specifying delta is integral to determining the BCV. It follows immediately that, since parity is rejected if the computed value of the test statistic "exceeds" the BCV, the value chosen for δ can determine the outcome of the test.

While the statistician may not be in a position to accurately specify δ, he or she is certainly able to evaluate the impact of choosing a particular δ on parity testing. Before turning to this question, however, let us examine briefly the ability of "telephony experts" to specify δ. In the past, BST "experts" have suggested that δ should equal 1; more recently (in the Florida Strawman proposal) a value of 0.5 has been put forward. No explanation has been offered as to how these numbers were derived. The following scenario is not out of the question: One day the chief ILEC negotiator phones one of his engineers and asks, "Hey Joe, suppose our average service provision was about one standard deviation better than what we provide the CLECs on average. Would that difference be competitively significant?" Joe thinks for a minute and responds, "Yeah, it probably would be, but let me check with Bill to see what he thinks. Hey, Bill…" To make a long story short, let's suppose that Bill and whoever else he consults concur. The value of δ has now been established, in the ILEC's mind, as 1. Admittedly, there is no real evidence to support this conjecture; but equally, there is no real evidence refuting it, either. That is one of the problems, ILECs provide no evidence from their "telephony experts" at all.

Charitably, the ILEC may simply have asked its experts the wrong question. It is probably true that selecting $\delta=1$, produces a means difference, $1 \bullet \sigma$, that is competitively significant. But the important question is whether this is the least possible means difference that would be competitively significant. If one is willing to accept values of δ that lead to inframarginal differences in competitive significance, then there is an infinity of equally legitimate values that δ could take on. For example, if $\delta=1$ results in a competitively significant means difference ($1 \bullet \sigma$), then so would values of $\delta=2,3,4, \ldots$, because they would lead to larger means differences than that given by $\delta=1$ (i.e., $2 \bullet \sigma$,

$3 \bullet \sigma$, $4 \bullet \sigma$, ...). Thus, specifying inframarginal values for δ becomes completely arbitrary, so that such values can contribute nothing to the solution of parity testing problems. **The real question is how *small* can δ be made and the resulting means difference be competitively significant.** Is it possible for means differences resulting from δ values of 0.5, 0.25, or 0.1 to be competitively significant differences? **It is the value of δ that leads to the marginally competitively significant means difference that we require, because it is the only unique, unambiguous, meaningful value to assign to δ if competitive significance is to be the criterion by which we determine Type II error.** *For this reason, establishing the δ that leads to marginally competitively significant means difference should be the subject of considerable research on the part of economists and statisticians as well as engineers and other "telephony experts."* The CLECs are aware of no models that have been estimated, no experiments that have been conducted by the ILECs. Indeed, the ILEC is typically in a uniquely poor position to conduct tests and experiments to establish the extent of marginally competitively significant differences in the provision of local telephone service because, generally speaking, it does not "compete" in local markets. In fact, a sound argument can be made that it is not possible at this time to accurately establish such values, because up to now, local telephone markets in the U.S. have not seen vigorous competition between the CLECs and the ILEC. Until such competition is the rule of the day, determining "competitive significance" can be based on nothing but conjecture.

III. The Statistical Consequences of Choosing a δ That is "Too Large"

Now consider the impact on parity testing of the ILEC's choice of $\delta=1$ rather than some, more appropriate, smaller number. The answer, in a nutshell, is this: the larger δ, the more extensive is the ILEC's carte blanche to thwart local competition. The rationale is as follows: (i) Larger values of δ indicate larger differences in SQM means. (ii) The larger the means difference, the less likely the commission of a type II error, i.e., the lower is β. (iii) Smaller values of β require smaller values of α to balance the two risks. (iv) Since α is not only the probability of committing a type I error but also the level of significance of the test, smaller values of α imply larger critical values of the test statistic. (v) Since larger means differences imply greater discrimination and since larger critical values of the test statistic make rejection of parity less likely, larger values of δ permit greater discrimination by the ILEC without its incurring a penalty. To see points (i)- (iv) more clearly, consider the Figure 1. The figure contains three sets of graphs with two graphs in each set. For each set, the upper graph can be considered as the distribution of ILEC sample means and the lower graph, as the distribution of CLEC sample means. The service being analyzed is assumed to be one in which larger numbers mean worse performance. Thus, in accordance with equations 1, the mean of the ILEC distribution is μ and the mean of the CLEC distribution is $\mu+\delta \bullet \sigma$. In the upper set of graphs, $\delta=1$, in the middle set, $\delta=0.5$, and in the lowest set, $\delta=0.25$.

Graphically, determining the balancing critical value is easy. The probability of a type I error is simply the area under the ILEC curve to the right of X* (ILEC sample means so large that they give the appearance of non-parity when parity is in fact the case), and the probability of a Type II error is the area under the CLEC curve to the left of X* (CLEC sample means so small that they give the appearance of parity when it is not truly the case). Determining the balancing critical value simply amounts to adjusting the dashed vertical line -- the one labeled BCV and the one that defines X* -- so as to equalize these two areas. Also note that even though the distributions are not normalized, it still follows that larger α (=β) areas imply smaller (in absolute value) critical values, and conversely.

Now consider the upper set of graphs which have been constructed under the hypothesis that δ=1. Here, the CLEC mean is a relatively large distance above the ILEC mean. Thus the BCV will determine α and β errors that are relatively small, indicating that the BCV itself will be relatively large in absolute value. Intuitively, since the CLEC mean is a relatively large distance above the ILEC mean, we are not very likely to commit a Type II error, that is, $\beta$is likely to be small. Consequently, α must also be small to equal β, and small α's correspond to large (in absolute value) critical values of the test statistic.

In comparison, consider the middle set of graphs. All factors are assumed to be the same as in the upper set except that now the CLEC mean is closer to the ILEC mean, δ=0.5 rather than δ=1. Relative to the first case, this increased proximity will lead to an increased β-risk and a BCV that cuts off larger areas in the tails of both distributions. Note that the larger α would correspond to a smaller (in absolute value) critical value of the test statistic.

Finally, note that the lowest set of graphs reinforces these notions. Again, everything is assumed to be the same as in the two earlier cases except that now the CLEC mean is closer still to the ILEC mean, δ=0.25. Again, because of this increased proximity, the α- and β-risks are higher and the resulting BCV lower (in absolute value) than in the previous cases.

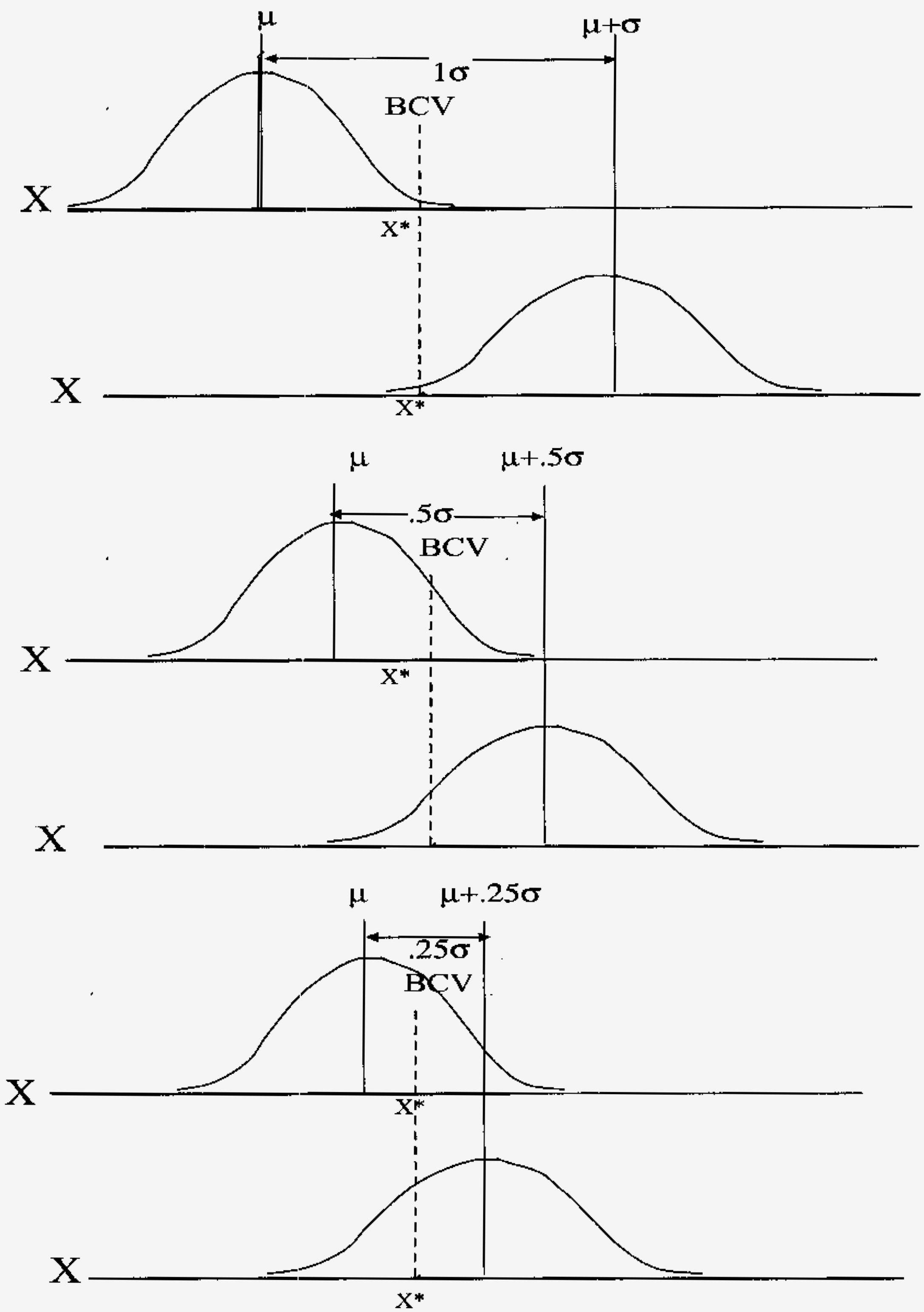
This analysis clearly demonstrates that, in general, the larger δ, the larger the critical value of the test statistic associated with the rejection of parity, *ceteris paribus.* Based on this result, it would not be difficult to accept a value of δ of 1 if the α- and β-risks were of a reasonable size; i.e., if the critical values of the test statistic were of reasonable magnitudes. Unfortunately, this is not the case for δ=1, nor even for δ=0.5. *The problem is that the AT&T/BST approach guarantees that, given δ, the α-risk will equal the β-risk, but it has nothing to say about the magnitude of risk at which they will be equal. As a result, many tests have critical values that balance risks, but at infinitesimal risk levels. In fact, these levels of significance are so small as to make a mockery of parity testing.*

Based on the hypothesis test defined in (1)

$$H_0: \mu_C = \mu_I; \sigma_C^2 = \sigma_I^2$$
$$H_A: \mu_C = \mu_I + \delta \bullet \sigma_I; \sigma_I^2 = \lambda \sigma_C^2 \qquad (1')$$

Begin by assuming that λ=1. BST has suggested a simplified formula for approximating the BCV for the truncated Z statistic. (It should be noted that what BST calls the truncated Z is in fact a standard normal variate -- the truncated Z minus its mean and divided by its standard deviation -- so that its critical values are those of a traditional Z statistic).

FIGURE 1



$$BCV = \frac{-\delta}{2\sqrt{\frac{1}{n_C}+\frac{1}{n_I}}} \qquad (3)$$

Let us begin by assuming that δ=1, and let us assume that the ILEC sample size is sufficiently large so that the term ($1/n_I$) in the denominator of (3) can be taken to be zero. Under these assumptions, the BCV depends only on δ and the CLEC sample size. Consider some typical CLEC sample size values, and note the implied values of BCV and the concomitant level of significance α (=β):

$n_C = 50 \Rightarrow BCV = -3.54 \Rightarrow \alpha = \beta = .0002$
$n_C = 100 \Rightarrow BCV = -5.00 \Rightarrow \alpha = \beta = .0000003$
$n_C = 300 \Rightarrow BCV = -8.66 \Rightarrow \alpha = \beta = 2.3*10^{-16}$
$n_C = 500 \Rightarrow BCV = -11.18 \Rightarrow \alpha = \beta = 2.5*10^{-28}$
$n_C = 1000 \Rightarrow BCV = -15.81 \Rightarrow \alpha = \beta = 1.3*10^{-54}$

It should be clear that, for very reasonable CLEC sample sizes, when δ=1, the AT&T/BST BCV approach yields unacceptably large (in absolute value) critical values and unacceptably small levels of significance. Put into perspective, the FCC has suggested that α=0.05 (CV=-1.645) is a reasonable significance level to undertake statistical tests of parity. Some ILEC proposals have suggested α=0.025 (CV=-1.96) or even α=0.01 (CV=-2.365). But no *bona fide* statistician could honestly recommend that it would be reasonable to conduct a simple means difference test at anything smaller than the α=0.01 level of significance -- that is, until now. By requiring δ=1, BST has implicitly required that the level of significance be 1/50th of the minimum acceptable level and 1/250th of an appropriate level -- in their best case scenario (n_C = 50). For more reasonable sample sizes, the implications are even more outrageous. And these results are not an artifact of the simplifying assumptions used in the above analysis. BST analyzed 84 parity tests on two SQMs using April 1999 data for the state of Louisiana, with δ=1. They report a minimum BCV of -73 (!) and a median BCV of -3.74, implying that half of the tests were undertaken at a level of significance less than .00009. Indeed, roughly 3/4th s of the tests were undertaken at less than the recommended .05 level of significance. These results indicate that, regardless of the opinion of the "telephony experts," the idea that δ=1 can be rejected based on its statistical implications alone.

These same conclusions also obtain in the case of δ=0.5, although to a lesser degree. Recall that this is the value of δ that BST has put forward in their Florida "Strawman" proposal. If we repeat the above experiment with δ=0.5, we find the following:

$n_C = 50 \Rightarrow BCV = -1.77 \Rightarrow \alpha = \beta = .038$
$n_C = 100 \Rightarrow BCV = -2.50 \Rightarrow \alpha = \beta = .0062$
$n_C = 300 \Rightarrow BCV = -4.33 \Rightarrow \alpha = \beta = .000007$
$n_C = 500 \Rightarrow BCV = -5.59 \Rightarrow \alpha = \beta = .00000001$
$n_C = 1000 \Rightarrow BCV = -7.91 \Rightarrow \alpha = \beta = 1.3*10^{-13}$

Again, except for the n_C=50 case, all significance levels are less than the minimum acceptable level, and even for the n_C=50 case, the significance level is less than the recommended .05 level. Thus, for the reasons mentioned above, δ=0.5 must be rejected on the grounds of its statistical implications as too big. (We acknowledge that these numbers do not dovetail with those in examples found in Appendix D of the BST proposal. They do, however, dovetail with the numbers we compute using that same data but appropriate, exact, formulae from other sources.)

Finally, prior to his retirement, AT&T's Colin Mallows recommended a value of 0.25 for δ. Replicating the above experiment for δ=0.25 yields

$n_C = 50 \Rightarrow BCV = -0.88 \Rightarrow \alpha = \beta = .19$
$n_C = 100 \Rightarrow BCV = -1.25 \Rightarrow \alpha = \beta = .106$
$n_C = 300 \Rightarrow BCV = -2.16 \Rightarrow \alpha = \beta = .015$
$n_C = 500 \Rightarrow BCV = -2.80 \Rightarrow \alpha = \beta = .0026$
$n_C = 1000 \Rightarrow BCV = -3.95 \Rightarrow \alpha = \beta = .00004$

Judged by the implied level of significance of the test, these results are considerably more credible than the two previous cases. Still, for instances where n_C>100, the levels of significance are just too low. This inference is particularly important since both AT&T and BST plans recommend aggregating the test statistics up through many deep testing categories before comparing them to the BCV, so that large CLEC sample sizes are to be expected. (To illustrate, the relevant sample sizes in the previously mentioned BST examples are in excess of n_C=300.)

IV. Implications for Parity Testing, Performance Appraisal, and the Prospects for Operationalizing Equal Risk

The practical import the above statistical results concerning parity testing should be obvious: The larger the value of δ, the greater the means difference, i.e., the greater the extent of discrimination against the CLEC, permitted the ILEC before it is subject to a penalty payment. An example will illustrate: The ILEC owes a penalty when the computed value of the test statistic exceeds the BCV. For simplicity, assume the test statistic is the LCUG Z and that $n_{ILEC} \to \infty$. Thus a penalty is owed if

$$\frac{\overline{X}_{CLEC} - \overline{X}_{ILEC}}{\sigma_{ILEC}\sqrt{\frac{1}{n_{CLEC}}}} \geq BCV \qquad (3)$$

Substituting equation (2) for BCV and rearranging terms, a penalty will be owed if

$$\overline{X}_{CLEC} \geq \overline{X}_{ILEC} + 0.5 \cdot \delta \cdot \sigma_{ILEC} \qquad (4)$$

Now suppose the ILEC mean repair interval is, say 3 days with a standard deviation of 8. If $\delta = 1$, the CLEC mean repair interval would have to be ***more than 7 days*** (as compared to the ILEC's 3 days) before the ILEC would owe a penalty. Indeed, if $\delta = 0.5$, as suggested in the Florida Strawman, the CLEC mean repair interval would have to be ***more than 5 days*** (as compared to the ILEC's 3 days) before the ILEC would owe a penalty. Interestingly, if $\delta = 0.15$, the implied means difference would be 0.6 days, about the same as that implied by the critical Z value of 1.645 (with $n_{CLEC} = 400$) suggested by the FCC (0.67 days).

This example should make it clear why ILECs want large values of δ and CLECs want small values of δ. It should also make it clear why δ has become such an important bargaining chip in 271 negotiations. *It is impossible to emphasize strongly enough how* *regrettable this outcome is.* The value of δ is not something to be bargained over any more than the value of π is something to be voted on. As pointed out in section II, δ is the difference between mean CLEC and ILEC performance levels, measured in units of the ILEC standard deviation, that would be marginally competitively significant. Ideally, its value for many different SQMs would be the subject of serious study by statisticians, economists, engineers, and industry experts. To make δ subject to negotiation is to destroy the logical underpinnings of parity testing and performance appraisal – to make these underpinnings rest on the relative bargaining power of the participants rather than statistical science. Yet this result is as inevitable as night following day. Because we have not seen at the local level the kind of vigorous competition among providers that would allow an appropriate calculation of δ, the only methods available for specifying δ are conjecture and negotiation, hopefully tempered with a little statistical sanity.

Problems arising from the acceptance or rejection of parity are not the only practical problems arising from attempts to apply the BCV approach. Such problems are magnified when the BCV approach enters into the determination of the magnitude of penalties. Consider for example the penalty structure in the Florida Strawman proposal. In that plan, the computed value of the (truncated) Z (call it Z*) and the BCV (the parity gap) is divided by 4 and the resulting percentage (called the "volume proportion," it cannot be >=1) which is then multiplied by the number of impacted CLECs to determine the "Affected Volume." This number multiplied by the per-occurrence penalty determines the payment to the CLEC for discriminatory service. Since penalties are owed only when Z*>BCV, increases in δ increase the BCV, which decreases the parity gap (for a given Z*), which decreases the volume proportion, which decreases the affected volume (for a given number of impacted CLECs), and hence lowers the penalty payment -- or the likelihood of a penalty being owed. *This means that by manipulating δ, the ILEC can manipulate penalty payments in such a way as to circumvent the intent of even the most adroit state oversight agencies.* Other plans involving δ and the BCV (e.g., AT&T's), while more reasonable, have similar potential of not reflecting the harm of disparity in a real world environment.. CLECs like WorldCom have agreed in joint CLEC remedy proposals to .25 as a generous trial as a BCV individual CLEC results. But WorldCom is becoming increasingly alarmed, as it should well be, that regulators are splitting the difference between ILEC and CLEC proposals for BCV's without any considered analysis of the impact of this "guess" of competitive significance on the marketplace.

V. Can Equal Risk Be Made Operational?

In principle, the BCV approach is indeed a beautiful dream. It eliminates the problem of random variation, and it reduces to zero the expected value of inappropriate penalty payments. Unfortunately, the crucial parameter δ cannot be unambiguously determined, there is an incentive on the part of the ILEC (CLECs) to inflate (deflate) δ, and making the value of δ a bargaining chip destroys the statistical legitimacy of parity testing and performance appraisal. The ILEC cannot be expected to make an enlightened choice of δ because it has scant experience with competition. The CLECs cannot be expected to make an enlightened choice of δ because they have limited experience in terms of contracting with the ILEC and with providing services in the local market. Since the kind of research needed to obtain an enlightened choice of δ is not possible at the present time, and since conjecture and negotiation clearly incorporate incentives to game the system, some CLECs (in particular, WorldCom) ***worry that a one-size-fits-all BCV can ever be made operational.***

For a moment, let us suspend disbelief and suppose that a BCV -- even with all its potential pitfalls -- is adopted. Would this be a good thing for the CLECs, the ILECs, the state regulatory agencies, or society as a whole? Even ignoring all of the problems brought to light up to now, the answer is still, "No!" Here is why: Suppose that in spite of all the impediments that the various BCV plans place before it, competition still develops. Increased competition implies larger CLEC orders, and larger CLEC orders imply lower probabilities if type II errors, *ceteris paribus*. But lower values of β imply lower balancing values of α, which in turn imply larger BCVs. Consequently, under the BCV approach, increased competition will make it less likely to judge a given means disparity as indicative of discrimination. This consequence is clearly unacceptable. A given difference in the quality of services provided by the ILEC to its own customers versus what it provides to those of the CLEC is either discriminatory or it is not. The extent of CLEC/ILEC competition should have nothing to do with this inference. For this reason, the long run acceptability of BCVs is even more uncertain than its short run acceptability.

It remains but to conclude that implementing a BCV approach is a risky strategy indeed. The CLECs support AT&T's proposal of a BCV approach only to the extent that it's proposed value of $\delta = 0.25$ is taken to be a ***maximum*** acceptable trial value of that parameter for individual CLEC results. This position is based on statistical sanity; conjecture, bargaining, or further alterations to increase the BCV are not acceptable. If state regulatory commissions find this position too intransigent, then some method other than the BCV approach must be found to deal with random variation and competitive significance.