ORIGINAL

MOYLE, FLANIGAN, KATZ, RAYMOND, WHITE & KRASKER, P.A.

ATTORNEYS AT LAW

RECEIVED-FPSC

06 SEP 15 PM 3:26

COMMISSION CLERK

The Perkins House
118 North Gadsden Street
Tallahassee, Florida 32301

Telephone: (850) 681-3828
Facsimile: (850) 681-8788

Wellington Office
(561) 227-1560
West Palm Beach Office
(561) 659-7500

Vicki Gordon Kaufman
E-mail: vkaufman@moylelaw.com

September 15, 2006
Via Hand Delivery

Ms. Blanca Bayo
Florida Public Service Commission
2450 Shumard Oak Boulevard
Tallahassee, FL 32399

Re: CLEC Response to Action Items
Docket No. 000121A

Dear Ms. Bayo:

Attached please find the original and two copies of the CLEC Coalition's responses to action items from the August 21, 2006 Six Month Review workshop for Docket 000121A. The attachments consist of:

Attachment 1:	Language from an ICA regarding hours of operation
Attachment 2:	Direct and rebuttal testimony of Collin Mallows CLEC Reply Comments (Issue 1) filed in Louisiana Statistician's Report from Louisiana
Attachment 3:	Response will be provided as soon as possible.
Attachment 4:	Additional language for Section 4.6.1 of SEEM plan
Attachment 5:	Comments regarding the feasibility of using parity test results at the CLEC aggregate level, etc.

CMP
COM
CTR
ECR
GCL
OPC
RCA
SCR
SGA
SEC 1
OTH Cover Ltr

Sincerely,

Vicki Gordon Kaufman

VGK/pg
Enclosures
cc: Parties of Record

RECEIVED & FILED
FPSC-BUREAU OF RECORDS

DOCUMENT NUMBER-DATE
08507 SEP 15 06
FPSC-COMMISSION CLERK

Attachment 1

Language from an ICA regarding hours of operation

PRE-ORDERING, ORDERING, PROVISIONING, MAINTENANCE AND REPAIR

QUALITY OF PRE-ORDERING, ORDERING, PROVISIONING, MAINTENANCE AND REPAIR

1.1 BellSouth shall provide pre-ordering, ordering, provisioning, and maintenance and repair services to BTI that are equivalent to the pre-ordering, ordering, provisioning, and maintenance and repair services BellSouth provides to itself or any other CLEC, where technically feasible. The guidelines for pre-ordering, ordering, provisioning, and maintenance and repair are set forth in the various guides and business rules, as appropriate, and as they are amended from time to time during this Agreement. The guides and business rules are found at http://www.interconnection.bellsouth.com and are incorporated herein by reference.

1.2 For purposes of this Agreement, BellSouth's regular working hours for provisioning are defined as follows:

Monday – Friday – 8:00 a.m. – 5:00 p.m. (Excluding Holidays)
(Resale/UNE non-coordinated, coordinated orders and order coordinated-time specific)

Saturday - 8:00 a.m. – 5:00 p.m. (Excluding Holidays)
(Resale/UNE non-coordinated orders)

The above hours represent the hours, either Eastern or Central Time, of the location where the physical work is being performed.

To the extent BTI requests provisioning of service to be performed outside BellSouth's regular working hours, or the work so requested requires BellSouth's technicians or Project Manager to work outside of regular working hours, overtime billing charges shall apply. Notwithstanding the foregoing, if such work is performed outside of regular working hours by a BellSouth technician or Project Manager during his or her scheduled shift and BellSouth does not incur any overtime charges in performing the work on behalf of BTI, BellSouth will not assess BTI additional charges beyond the rates and charges specified in this Agreement.

2. ACCESS TO OPERATIONS SUPPORT SYSTEMS

2. BellSouth shall provide BTI access to operations support systems ("OSS") functions for pre-ordering, ordering, provisioning, maintenance and repair, and billing. BellSouth shall provide access to the OSS through manual and/or electronic interfaces as described in this Attachment. It is the sole responsibility of

2.10.1.3 BTI shall order Services and Elements as set forth in this Attachment 2 and BellSouth shall provide a Firm Order Confirmation ("FOC") (as that term and acronym are defined in Attachment 7, incorporated herein by this reference).

2.10.2 Ordering

2.10.2.1 BTI shall request Hot Cuts from BellSouth by delivering to BellSouth a valid Local Service Request ("LSR") using BellSouth's ordering interfaces described in Attachment 6 to this Agreement, incorporated herein by this reference. BTI may specify a Due Date or Frame Due Time, as defined below, at any time, including twenty-four (24) hours a day and seven (7) days a week. BTI shall specify whether its service order is to be provisioned by BellSouth as either: (a) Order Coordination ("OC"); or (b) Order Coordination—Time Specific ("OC-TS"). OC shall mean the type of service order used by BTI to request that BellSouth provision a Hot Cut on the particular calendar date as specified on the LSR and confirmed on the FOC as set forth in Section 2.10.2.3 below, at any time during that day, referred to in this Section as the "Due Date." OC-TS shall mean the type of service order used by BTI to request that BellSouth provision a Hot Cut on the particular day returned on the FOC as set forth in Section 2.10.2.3 below and at the particular time specified on the FOC, referred to in this Section as the "Frame Due Time." BTI shall pay the appropriate rate for either OC or OC-TS as set forth in Attachment 2. BTI will be billed and will pay overtime for conversions requested and occurring outside of BellSouth's normal hours of operation as defined in Section 2.10.2.2 below.

2.10.2.1 Until such time as BellSouth's systems can deliver the requested frame due time on the FOC as set forth above, BTI shall rely on the time requested on the LSR.

2.10.2.2 For purposes of this Section, BellSouth's normal hours of operation for personnel performing physical wire work are defined as follows:

2.10.2.2.1 Monday – Friday: 8:00 a.m. –5 :00 p.m. (Excluding Holidays) (Resale/UNE non-coordinated, coordinated orders and order coordination time specific)

2.10.2.2.2 Saturday: 8:00 a.m. – 5:00 p.m. (Excluding Holidays) (Resale/UNE non-coordinated orders)

2.10.2.2.3 The above hours are defined as the time of day where the work is being performed.

2.10.2.2.4 Normal hours of operation for the various BellSouth centers supporting ordering, provisioning and maintenance are as set forth in Attachment 6 and incorporated herein by this reference. Normal hours of operation for the BellSouth centers providing BTI support will be equal to the hours of operation that BellSouth provisions services to its affiliates, end users, and other CLECs.

disconnect order, BellSouth shall use its best efforts to reconnect service within 24 hours.

4.2 BellSouth shall provide ITC^DeltaCom with a FOC for each Resale and UNE order. As of the date of this Agreement, the FOC includes purchase order number, telephone number, Local Service Request Number, the due date and Service Order number. Any changes to information included in the FOC shall be as determined by the EICCP.

4.3 BellSouth shall provision Resale Services and UNEs as prescribed in ITC^DeltaCom's service order requests. Access to status on such electronic orders of Resale services and UNEs shall be provided via the electronic interfaces utilized by ITC^DeltaCom. Status on manual orders shall be provided as mutually agreed by the Parties.

Order Status shall allow ITC^DeltaCom to check service order status, including Due Dates and Customer and Facility Due Date-Jeopardies.

BellSouth shall provide notice of a lack of facilities availability at parity (in terms of means and timing) to that BellSouth provides to itself, its Affiliates, or any other Telecommunications Carrier.

General Ordering/Provisioning Requirements

4.6.1 BellSouth shall provide a single point of contact ("SPOC") for the provisioning of Resale Services (LCSC) and provisioning of UNEs (UNE center) ordered by ITC^DeltaCom. For services and UNEs available electronically, preordering and ordering shall be available via an electronic interface seven (7) days a week, 24 hours a day less reasonable periods for maintenance and scheduled downtime. During provisioning of services to ITC^DeltaCom, support personnel will be available until the migration of the end user is complete. Provisioning services (LCSC and UNE Center) shall be provided during the same business hours that BellSouth provisions services to its own end users. All other ITC^DeltaCom requests for provision and installation services are considered outside of the normal hours of operation and may be performed subject to the application of additional charges.

4.6.2 BellSouth shall provide access to assistance for technical issues such as connectivity and passwords related to LENS, TAG and TAFI, and to the "EDI Central Group" for technical problems with EDI. Assistance will be available by telephone during normal business hours and through other contacts on nights, weekends and holidays.

4.6.3 BellSouth shall provide the following to ITC^DeltaCom:

4.6.3.1 Circuit Layout Record Card and Design Layout Records ("DLRs") for designed unbundled Network Elements;

5.7 BellSouth's Inside Wire Maintenance Service Plan may be made available for resale at rates, terms and conditions as set forth by BellSouth and without the wholesale discount.

5.8 BellSouth will provide customer record information to ITC^DeltaCom provided ITC^DeltaCom has either executed a blanket agency agreement or has the appropriate Letter(s) of Authorization. BellSouth shall provide customer record information via an electronic interface and in accordance with the provisions of Attachment 6.

5.9 Telephone numbers transmitted via any resold service feature are intended solely for the use of the end user of the feature. Resale of this information is prohibited.

6.0 Operations Support Systems Functions

6. BellSouth shall provide ITC^DeltaCom advance notice of changes to the prices, terms, and conditions for Resale in accordance with the provisions of Section 20.3 of the General Terms and Conditions. BellSouth provides electronic access to customer record information. Access is provided through the Local Exchange Navigation System (LENS), and the Telecommunications Access Gateway (TAG). Customer Record Information includes but is not limited to, customer specific information in CRIS and RSAG. ITC^DeltaCom agrees not to view, copy or otherwise obtain access to the customer record information of any customer without that customer's permission and only in accordance with applicable federal and state regulations.

6.2 As provided in Section 3 of the General Terms and Conditions and Attachment 6, BellSouth shall provide ITC^DeltaCom, at its request, non-discriminatory access to BellSouth's OSS functions for pre-ordering, ordering, provisioning, maintenance and repair, and billing. Such OSS functions shall be equal in quality and provisioned with the same timeliness as provided by BellSouth to itself or to any Subsidiary, Affiliate or any other Telecommunications Carrier to which BellSouth provides the OSS functions.

6.3 Charges for use of OSS shall be as set forth in Exhibit A of this Attachment and in Attachment of this Agreement.

7.0 Maintenance of Services

7. ITC^DeltaCom will adopt and adhere to the standards contained in the applicable BellSouth Work Center Interface Agreement regarding maintenance and installation of service.

7.2 Services resold under BellSouth's Tariffs and facilities and equipment provided by BellSouth shall be maintained by BellSouth

7.3 ITC^DeltaCom or its end users may not rearrange, move, disconnect, remove or attempt to repair any facilities owned by BellSouth, other than by connection or disconnection to any interface means used, except with the written consent of BellSouth.

7.4 ITC^DeltaCom accepts responsibility to notify BellSouth of situations that arise that may result in a service problem.

7.5 ITC^DeltaCom will be BellSouth's single point of contact for all repair calls on behalf of ITC^DeltaCom's end users. The parties agree to promptly provide one another with toll-free contact numbers for such purposes.

7.6 ITC^DeltaCom will contact the appropriate repair centers in accordance with reasonable procedures established by BellSouth

Parties continue beyond the expiration date of this Agreement to negotiate the local interconnection arrangements without Commission intervention, the terms, conditions and prices ultimately ordered by the Commission, or negotiated by the Parties, will be effective retroactive to the day following the expiration date of this Agreement. Until the Subsequent Agreement becomes effective, the Parties shall continue to exchange traffic pursuant to the terms and conditions of this Agreement.

2. Ordering Procedures

2.1 Detailed procedures for ordering and provisioning BellSouth services are set forth in BellSouth's Local Interconnection and Facility Based Ordering Guide, Resale Ordering Guide, and as set forth in Attachment 6 of this Agreement, as appropriate. 2.2 BellSouth has developed electronic systems for placing most resale and some UNE orders. BellSouth has also developed electronic systems for accessing data needed to place orders including valid address, available services and features, available telephone numbers, due date estimation on pre-order and calculation on firm order, and customer service records where applicable. Charges for OSS shall be as set forth in Attachment 1, Exhibit A, Attachment 2 and in Attachment 11 of this Agreement.

3. Parity

3.1 The services and service provisioning that BellSouth provides ITC^DeltaCom for resale will be at least equal in quality to that provided to BellSouth, or any BellSouth subsidiary, affiliate or end user. In connection with resale, BellSouth will provide ITC^DeltaCom with pre-ordering, ordering, maintenance and trouble reporting, and daily usage data functionality that will enable ITC^DeltaCom to provide equivalent levels of customer service to their local exchange customers as BellSouth provides to its own end users.

3.2 BellSouth shall also provide ITC^DeltaCom with unbundled network elements, and access to those elements. The quality of an unbundled network element, as well as the quality of the access to such unbundled network element, that BellSouth provides to ITC^DeltaCom shall be at least equal in quality to that which BellSouth provides to itself or to any BellSouth subsidiary, affiliate or other CLEC. The terms and conditions pursuant to which BellSouth provides access to unbundled network elements, including but not limited to the time within which BellSouth provisions such access to unbundled network elements, shall, at a minimum, be no less favorable to ITC^DeltaCom than the terms and conditions under which BellSouth provisions such elements to itself. Consistent with all applicable rules and regulations, BellSouth shall provide ITC^DeltaCom with pre-ordering, ordering, provisioning, maintenance and repair, and billing functionality at least equal to that which BellSouth provides for its own retail services.

Attachment 2

Direct and rebuttal testimony of
Collin Mallows
CLEC Reply Comments (Issue 1)
filed in Louisiana
Statistician's Report from Louisiana

BEFORE THE GEORGIA PUBLIC SERVICE COMMISSION

DIRECT TESTIMONY OF COLIN MALLOWS, PH.D.

ON BEHALF OF

THE CLEC COALITION

DOCKET NO. 7892-U

JUNE 20, 2000

Q. PLEASE STATE YOUR NAME AND BUSINESS ADDRESS.

A. My name is Colin Mallows. My business address is AT&T Labs, 180 Park Avenue, Florham Park, New Jersey 07932-0971.

Q. PLEASE DESCRIBE YOUR EDUCATIONAL AND PROFESSIONAL BACKGROUND.

A. I received a Ph.D. in Statistics from the University of London (England) and have been employed by AT&T since 1960 at Bell Labs and AT&T Labs. I was a Lecturer at the University of London from 1955-1960. I also was an Adjunct Associate Professor at Columbia University from 1960-64. I was a Department Head at Bell Labs from 1969-1986. I have authored over 100 papers on statistics that have been published in many professional journals. I was the statistician who represented AT&T in the creation of the original Local Competition Users Group ("LCUG") 1.0 paper that was published in February, 1998. I have represented AT&T and other CLECs in several regulatory proceedings concerning the appropriate

statistical methodology to use in an effective performance measures methodology. I have met with the FCC on this issue and have participated in state regulatory workshops and meetings in Louisiana, California, Nevada, Texas, and New York.

Q. PLEASE SUMMARIZE YOUR TESTIMONY.

A. My testimony describes the statistical methodology for use in comparing BellSouth's performance for itself and its affiliates to the performance it provides to CLECs. In addition, I describe the results of work I have done with statisticians from Ernst & Young who have worked on behalf of BellSouth, to resolve some issues identified in the ongoing performance measures workshop in Louisiana.[1]

Q. WHY IS A STATISTICAL TEST NECESSARY?

A. Once an appropriate basis for comparison has been established, a tool is needed to determine quantitatively whether BellSouth has provided nondiscriminatory treatment. Merely reporting averages of performance measurements alone, without further analysis, does not indicate whether differences in performance results reflect discrimination. In fact, averages may even mask discrimination. The FCC supported the use of statistical comparisons in its recent Bell Atlantic Order. *See* In the Matter of Application of Bell Atlantic for Provision of In-Region InterLATA Services

[1] Louisiana Public Service Commission, Docket No. 22252-C, In re: BellSouth Telecommunications, Inc, Service Quality Measurements.

In New York, CC Docket No. 99-295 (December 23, 1999), Appendix B, where it stated:

> "When making a parity comparison, statistical analysis is a useful tool to take into account random variations in the metrics. In the Second BellSouth Louisiana Order, we encouraged BOCs to submit data allowing us to determine if any detected difference between the wholesale and retail metrics is statistically significant."

A statistical test should be applied only to those measures for which there is a retail analog. Regardless of which parity measure is under consideration, there must be a pre-established comparison process to assure that the levels of performance both for an individual CLEC, and the CLECs as a group, are at least equal in quality to BellSouth's performance for its own retail service operation or that of BellSouth's affiliate. This comparison process for parity measures is completed through the use of a statistical test. This Commission should require BellSouth to apply statistical testing to all performance results for parity measures and report the conclusions. BellSouth also should be required to provide sufficient underlying detail for benchmark measures to permit CLECs to determine how many individual data points failed to achieve the identified benchmark level of performance.

Q. WHAT STATISTICAL METHODOLOGY DO THE CLECS RECOMMEND?

A. The statistical methodology recommended by the CLECs is based on use of the modified z statistic. The methodology is described in a paper attached to this testimony as Exhibit A.[2] For each parity submeasurement (a disaggregated measure), the difference between BellSouth's performance for its retail operation or that of its affiliates and the performance it provides for a given CLEC is converted to a z-value (the modified z statistic). Out-of-parity performance occurs when the z-value exceeds an agreed upon critical value. In addition, the "pooled Z" formula is recommended for sample sizes with less than 30 data points The statistical methodology requires that a critical value (depending on the numbers of observations) be chosen; values of Z that are less than the critical value are taken to be indications of discrimination. In the Louisiana proceeding, AT&T and BellSouth have agreed on a methodology (based on a "balancing" concept) for determining the critical value. However, they have not yet been able to agree on the appropriate value of the number "delta" that is required in this calculation. In the absence of such agreement, AT&T recommends that the critical value be taken as -1.04, since this value is an approximation to what the full "balancing" calculation would give.

[2] *See* Exhibit A, "Statistical Tests for Local Service Parity", Version 1.0, February 6, 1998, Local Competition Users Group.

Q. IS THE STATISTICAL METHODOLOGY RECOMMENDED BY CLECS COMPLETE IN ITS DEVELOPMENT?

A. Yes. As previously stated, the appropriateness of the methodology has also been validated as part of the Louisiana proceeding.

Q. DESCRIBE THE STATISTICAL METHODOLOGY ISSUES THAT WERE CONSIDERED IN THE ANALYSIS OF DATA IN THE ONGOING PERFORMANCE MEASURES PROCEEDING IN LOUISIANA.

A. The Louisiana Public Service Commission included language in an August 31, 1998 order in Docket No. U-22252-C requiring BellSouth to give CLECs access to raw data that underlies BellSouth's reports.[3] In that proceeding, AT&T entered into a protective agreement with BellSouth so that AT&T's statistician could receive at least some of BellSouth's performance data and work with it for analyzing the proper working of the statistical test.[4] The ability to look at the data and analyze it is critical to determining the appropriate statistical test. One cannot be assured that the data characteristics are properly accounted for in the statistical methodology unless one can observe the data and how it behaves over time. The Louisiana Public Service Commission's order provided the

[3] Order, In re: BellSouth Telecommunications, Inc., Service Quality Performance Measurements, Docket No. U-22252, Subdocket C, August 31, 1998.

[4] BellSouth provided some of its raw data associated with four measures it includes in its SQM. The measures for which AT&T's statistician received some raw data were: Order Completion Interval, Maintenance Average Duration, Missed Repair Appointments, and Missed Installation Appointments.

opportunity to actually see raw data and, thereby, confirm and refine the statistical methodology.

Several issues were considered in the CLECs' analysis of the data, including whether the modified z statistic, as explained in the LCUG paper, was feasible (considering this was the first opportunity to apply the modified z statistic to actual data), whether the modified z statistic properly handled small sample sizes, whether the results of the modified z statistic methodology differed from the results BellSouth obtained using its "jackknife" method (a test statistic proposed by BellSouth in LA), and if those results differed, why they did. The original LCUG proposal did not address the aggregation issue, but the AT&T statistician proposed a way of aggregating modified Z from the cell level for comparison with BellSouth's then proposed "jackknife" method.

Q. WHAT CONCLUSIONS WERE REACHED AFTER ANALYZING THE DATA?

A. I analyzed the raw data which confirmed the following: (1) the modified z statistic is an effective component of the methodology for parity determinations; (2) there were some issues to resolve with handling small sample sizes; and (3) the method based on LCUG's modified z statistic and BellSouth's "jackknife" method produced different results.

The data provided a means to confirm why the two approaches produced different results. The "jackknife" method proved to be less effective in detecting discriminatory performance. In the course of the Louisiana performance measures proceeding, AT&T's and BellSouth's statisticians were requested by the Louisiana Public Service Commission Staff to collaborate and fix the deficiencies of the "jackknife" method, which inappropriately aggregated several cell results into a single cell for determining a test statistic. Fixing BellSouth's jackknife deficiencies resulted in the development of the truncated z statistic.

Q. WHAT WERE THE ADDITIONAL PRIMARY CONCLUSIONS THAT WERE REACHED AFTER ANALYZING THE DATA?

A. BellSouth's statisticians and I concluded through the work they did in the Louisiana proceeding that in performing permutation calculations for small samples, it is not necessary to use the LCUG formula. That formula can be replaced by the simpler and faster "pooled Z" formula. The statisticians also concluded that aggregation of results from many small cells into a single overall statistic raises several new problems that had not been addressed in the LCUG paper given that results for modified z were assessed at the submeasure level without considering the need for aggregation of several cells into a single overall statistic. The statisticians also concluded that the method they developed for balancing the critical value is an efficient and quantitative means of establishing a critical value.

Q WHY IS THE CRITICAL VALUE IMPORTANT?

A. The critical value is the value which determines whether parity or out-of-parity exists.

Q. WHAT CRITICAL VALUE DO THE CLECS RECOMMEND?

A. The CLECs recommend as the critical value –1.04 which is only an approximation to what the "balancing" approach would give.

Q. WHAT IS A TYPE 1 ERROR?

A. This is an error that may occur due to random variation that indicates that BellSouth is favoring its retail operations, when in fact, it is not.

Q. WHAT IS A TYPE 2 ERROR?

A. This is an error that may occur due to random variation that indicates that BellSouth is not favoring its retail operations, when in fact, it is.

Q. IS THE DEVELOPMENT OF THE BALANCING CRITICAL VALUE COMPLETE?

A. No, AT&T's and BellSouth's statisticians agree on the principles underlying the approach to balancing Type 1 and Type 2 errors. However, a decision must be made to determine the appropriate value of the "delta" parameter which specifies the degree of non-compliance that is judged to

be serious. The Balancing Critical Value development is incomplete until the value of parameter "delta" is determined.

Q. IS THE DECISION ABOUT THE VALUE OF THE PARAMETER DELTA REQUIRED BEFORE IMPLEMENTING THE CLECS' CURRENT STATISTICAL METHODOLOGY?

A. No. The CLECs use –1.04 as the critical value. It is at –1.04 that the probability of Type 1 or Type2 Errors are approximately balanced.

Q. WHY HAS THE DETERMINATION OF THE "DELTA" PARAMETER NOT BEEN RESOLVED?

A. There is agreement that balancing should be done and on the formulas to be used. The only unresolved question is the value of the parameter "delta" which defines the degree of violation of parity at which the balancing should occur. Resolution of this question cannot be based solely on a technical analysis. Ideally, this decision should be based on business judgment, namely by consideration of how large a violation of parity must be before it is "important". The parameter "delta" measures the size of the violation. The larger the delta we choose, the smaller the balancing value of the errors. Once delta is chosen, the formula makes proper allowance for the effect of the sample size. BellSouth wants a large delta because this means a smaller Type 1 error and hence, larger Type 2 errors for all degrees of violations. When delta is large, the

balancing occurs at a more extreme degree of violation. The balancing occurs at a more extreme degree of violation, where the Type 2 error is smaller. The CLECs want a smaller delta because CLECs believe it is important to be able to detect a small but meaningful degree of violation, if it occurs.

Q. WHERE DO YOU BELIEVE THE DELTA FOR THE CRITICAL VALUE SHOULD BE SET?

A. The CLECs propose that this Commission adopt .25 as the parameter delta value. BellSouth proposed a delta equal to 1.0. To understand the implications of this choice, consider what it implies for how many customers receive bad service. Consider the level of service that BellSouth provides for the worst treated 1% of its own customers. Then, if we assume the observations are normally distributed, a violation with the delta equal to 1.0 means that 9.2% of CLEC customers will get service this bad, (*i.e.* the CLEC rate is more than nine times the BellSouth rate). Similar results will be obtained if we assume other distribution shapes. On the other hand, with delta set equal to 0.25, 1.8% of CLEC customers receive service this bad--still nearly twice the BellSouth rate but far better than the result with delta set equal to 1.0.

Consider a measure that is expressed as a percentage, for which BellSouth consistently achieves 90%. Then a delta equal to 1 corresponds to making the CLEC proportion 46.4%, while a delta equal to

0.25 corresponds to 81.3%. Similarly, if the BellSouth target is 99%, with a delta equal to 1 the CLEC alternative is 68.1%, while with a delta equal to 0.25 it is 95%. The delta equal to 1 alternatives are much too lenient by allowing far too many more CLEC customers to receive inferior service than ILEC customers.

Q. HOW DO THE AGREEMENTS ON HOW TO RESOLVE THE ISSUES IDENTIFIED DURING THE LOUISIANA PROCEEDING CHANGE THE BASIS OF THE STATISTICAL METHODOLOGY IN THE FEBRUARY 6, 1998 LCUG PAPER?

A. The agreements on the issues we discovered during the Louisiana performance measures proceeding do not represent a departure from support of the modified z statistic—the basis of the LCUG paper. We have simply enhanced the operation of the methodology so that it is more accurate once the Balancing Critical Value is complete in its definition and incorporated. The LCUG paper is now two years old. In addition, the LCUG paper was prepared when the Regional Bell Operating Companies refused to allow any review whatsoever of actual data. Only after the Louisiana Public Service Commission ordered access to the raw data did we get the opportunity to test the methodology with real data. After having the opportunity to actually see the raw data and how it behaves, we were able to enhance the LCUG methodology to handle small sample sizes using the permutation method. We developed the

"truncated z" methodology for aggregating small cells into a single overall statistic, and we developed the balancing method for choosing critical values.

Q. WHAT DO CLECS RECOMMEND THAT THIS COMMISSION ORDER CONCERNING THE STATISTICAL METHODOLOGY?

A. There are two things that should be included in the Commission's order. First, CLECs propose that this Commission order the modified z as a component of the statistical methodology. Second, CLECs propose that this Commission order the parameter delta value be set at 0.25 so that the companies may complete the development work concerning the Balancing Critical Value and then incorporate it into the CLECs' statistical methodology. In the absence of these recommended actions, the commission should order the use of –1.04 as the critical value that approximates the balanced result.

Q. WHEN THE DELTA VALUE IS ORDERED, WILL CLECS BE SATISFIED THAT THE RECOMMENDED STATISTICAL METHODOLOGY WILL ACCURATELY EVALUATE BELLSOUTH'S PERFORMANCE?

A. This is not a perfect statistical methodology. We do not believe a perfect methodology for this purpose can be created. However, this methodology will detect discrimination when the delta value for balancing the Type 1 and Type 2 errors is properly set. We expect to monitor how the

methodology works in "production mode", when very large amounts of data are being analyzed. AT&T's statistician will monitor how the methodology works after implementation and will make recommendations for improvements, if necessary, just as he did in the Louisiana proceeding when he had the opportunity to observe actual data.

Q. DOES THAT CONCLUDE YOUR TESTIMONY?

A. Yes.

BEFORE THE GEORGIA PUBLIC SERVICE COMMISSION

REBUTTAL TESTIMONY OF COLIN MALLOWS, PH.D.

ON BEHALF OF THE

CLEC COALITION

DOCKET NO. 7892-U

JUNE 27, 2000

Q. PLEASE STATE YOUR NAME AND BUSINESS ADDRESS.

A. My name is Colin Mallows. My business address is AT&T Labs, 180 Park Avenue, Florham Park, New Jersey 07932-0971.

Q. PLEASE DESCRIBE YOUR EDUCATIONAL AND PROFESSIONAL BACKGROUND.

A. I received a Ph.D. in Statistics from the University of London (England) and have been employed by AT&T since 1960 at Bell Labs and AT&T Labs. I was a Lecturer at the University of London from 1955-1960. I also was an Adjunct Associate Professor at Columbia University from 1960-64. I was a Department Head at Bell Labs from 1969-1986. I have authored over 100 papers on statistics that have been published in many professional journals. I was the statistician who represented AT&T in the creation of the original Local Competition Users Group ("LCUG") 1.0 paper that was published in February, 1998. I have represented AT&T and other CLECs in several regulatory proceedings concerning the appropriate statistical methodology to use in an effective performance measures methodology. I have met with the FCC on this issue and have

participated in state regulatory workshops and meetings in Louisiana, California, Nevada, Texas, and New York.

Q. WHAT IS THE PURPOSE OF YOUR REBUTTAL TESTIMONY?

A. The purpose of my rebuttal testimony is to address several statements made in the direct testimony of BellSouth's witness, Mr. Alphonso J.Varner. Specifically, I will show that BellSouth's proposed values for the parameter delta are excessive and unjustified and that a reasonable value is no greater than 0.25

Q. DR. MALLOWS, IN YOUR JUDGEMENT ARE VALUES FOR THE PARAMETER DELTA PROPOSED BY BELLSOUTH REASONABLE?

A. No. Mr. Varner's explanation of the definition of the parameter "delta" is accurate, but his proposal to use a delta of 1 and .5 lacks justification and therefore should be rejected by this Commission. Assuming that ILEC observations and CLEC observations are being generated by two different processes, the parameter "delta" is a measure of the difference between these two processes. It has the same interpretation, no matter how many observations are available. For ease of interpretation, it is helpful to think of delta as measuring the amount by which the ILEC distribution has to be shifted to make it match the CLEC distribution, measured in units of the ILEC standard deviation. Mr. Varner makes several arguments that are irrelevant, but nowhere does he give a business justification for setting delta as high as 1.0.

Q. WHAT CRITERIA SHOULD THIS COMMISSION USE IN DETERMINING AN APPROPRIATE VALUE FOR THE PARAMETER DELTA?

A. As Mr. Varner says, ideally the value of delta should be set by consideration of the impact of any given degree of shift on the customers' perception of service. To set up a test of parity performance, the Commission should choose a value of delta such that if this degree of violation is present, the Commission and others will have a very good chance of detecting it. To aid in making the judgement of the appropriate value of delta, it is helpful to look at pairs of distributions that differ by various amounts. This was done by AT&T in a response to an FCC inquiry on this very point. That response is attached to this testimony as AT&T CLM Rebuttal Exhibit No.1.

Q. PLEASE EXPLAIN YOUR REBUTTAL IN EXHIBIT NO.1.

A. This exhibit provides two examples of what various values of delta mean.

First, I looked at some data on Order Completion Interval given in a BellSouth report filed in Georgia in March 2000. This report did not give complete details, so I had to estimate the shape of the upper tail of the distribution of BellSouth observations. I then shifted this estimated distribution by amounts corresponding to delta = 0.25, 0.5, and 1.0, and prepared histograms representing what the corresponding CLEC distributions would be. Also, I gave a table showing what the effect of various values of delta would be for a counted variable, such as Missed Appointments. In both cases it appears that taking delta equal to = 1 represents an extremely large violation of parity. Delta = 0.25 is less extreme, but still represents a substantial degree of violation.

Q. DO YOU AGREE WITH MR. VARNER'S STATEMENT ON PAGE7, LINE 22, OF HIS TESTIMONY THAT A DELTA OF 1 "MEANS THAT INDIVIDUAL CLEC RESULTS WITHIN ONE (1) STANDARD DEVIATION OF BELLSOUTH'S RESULTS ARE NOT MATERIALLY DIFFERENT."?

A. No. There is no hard line between what is "materially different" and what is not; rather there is a gradual intensification of the degree of impact as the size of the shift (i.e. delta) increases. The fact that a shift of one standard deviation is regarded as meaningful for the purpose of setting up a fair test does not imply that all smaller shifts are not violations of parity treatment, which may deserve that penalties be assessed. We regard the value of delta as specifying a degree of violation so large that large penalties should be imposed. Smaller degrees of violation should incur smaller penalties.

Q. PLEASE COMMENT ON MR. VARNER'S STATEMENT REGARDING "THE NORMAL VARIANCE."

A. On page 8, line 19, Mr. Varner states that "Generally, three standard deviations is considered the normal variance above which things are considered to be out of line." This is a statement about statistical significance, not meaningful difference. Also, the factor "3" here is applied to the standard error of the observed average, not the standard deviation of the ILEC population. These two things differ by a factor of sqrt(n), where n is the number of observations in the average.

Mr. Varner appeals to the authority of Edwards Deming, who said: "[I]n practice, 3-sigma limits have been found to be the correct spacing..." The parameter delta was introduced in an attempt to avoid such a context-free rule, by basing the test on a judgement of materiality

rather than on some arbitrary convention. Note that in Deming's quote, the "3-sigma limit" is applied to the sample mean, and so will vary as a function of the number of observations. By contrast, delta is defined in a way that does not depend on sample sizes.

Q. IS THE CLEC COALITION IN THIS PROCEEDING PROPOSING A PARAMETER DELTA OF ZERO?

A. On page 9, line 3 and subsequently, Mr. Varner refers to a delta of zero. The Coalition has never suggested such a choice, which would make no sense because a delta of 0 means there is no difference between the ILEC and CLEC distributions.

Q. PLEASE SUMMARIZE YOUR REBUTTAL TESTIMONY.

A. In the absence of genuine engineering judgement, our study of pairs of distributions such as appear in AT&T's response to the FCC leads me to recommend that a delta of 0.25 should be chosen. Note that in the FCC example, the fraction of BellSouth customers whose OCI measurements exceed the BellSouth mean of 6.57 is estimated to be 25%. With a shift of magnitude delta = 0.25, we estimate that 39% of CLEC customers would exceed this value. With a shift to a delta of 1, 99.8% of CLEC customers would have observations above the BellSouth mean. In this example the degree of violation that is represented by a delta at 1 is very extreme indeed. Moreover, the parameter delta proposed by BellSouth is unjustified and unwarranted and should not be adopted by this Commission.

Q. DOES THAT CONCLUDE YOUR REBUTTAL TESTIMONY?

A. Yes.

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.

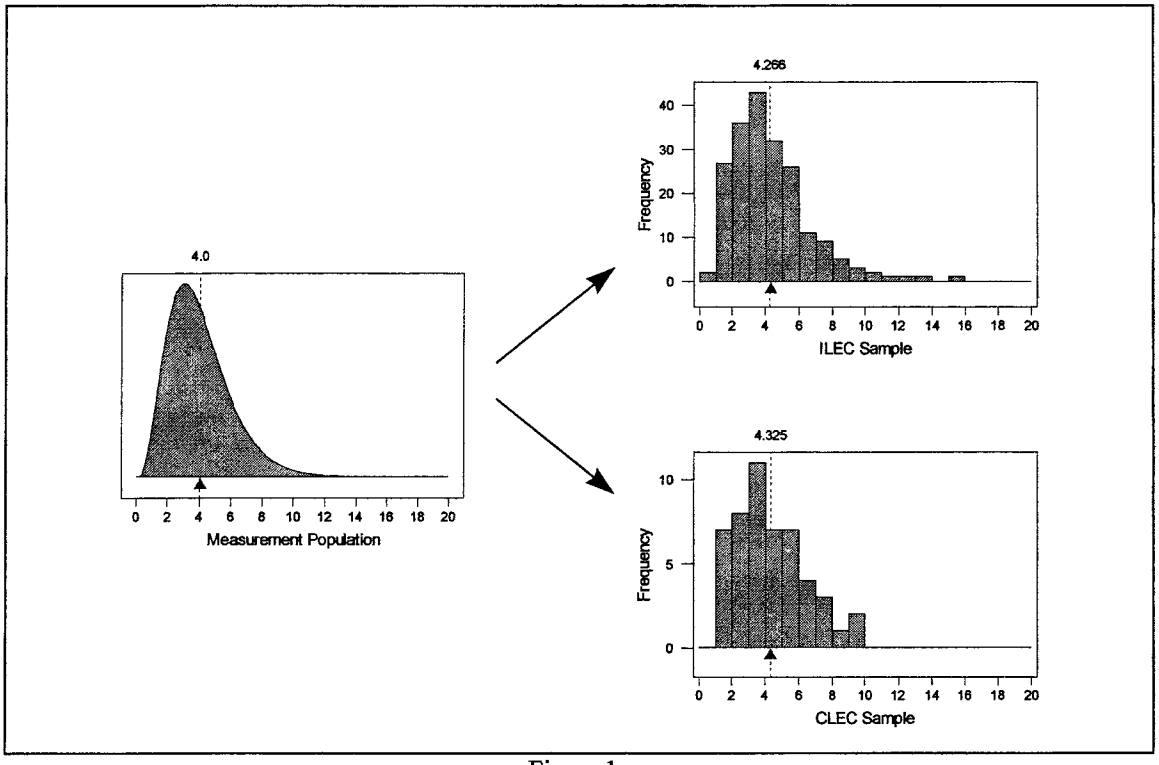


Figure 1.

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).

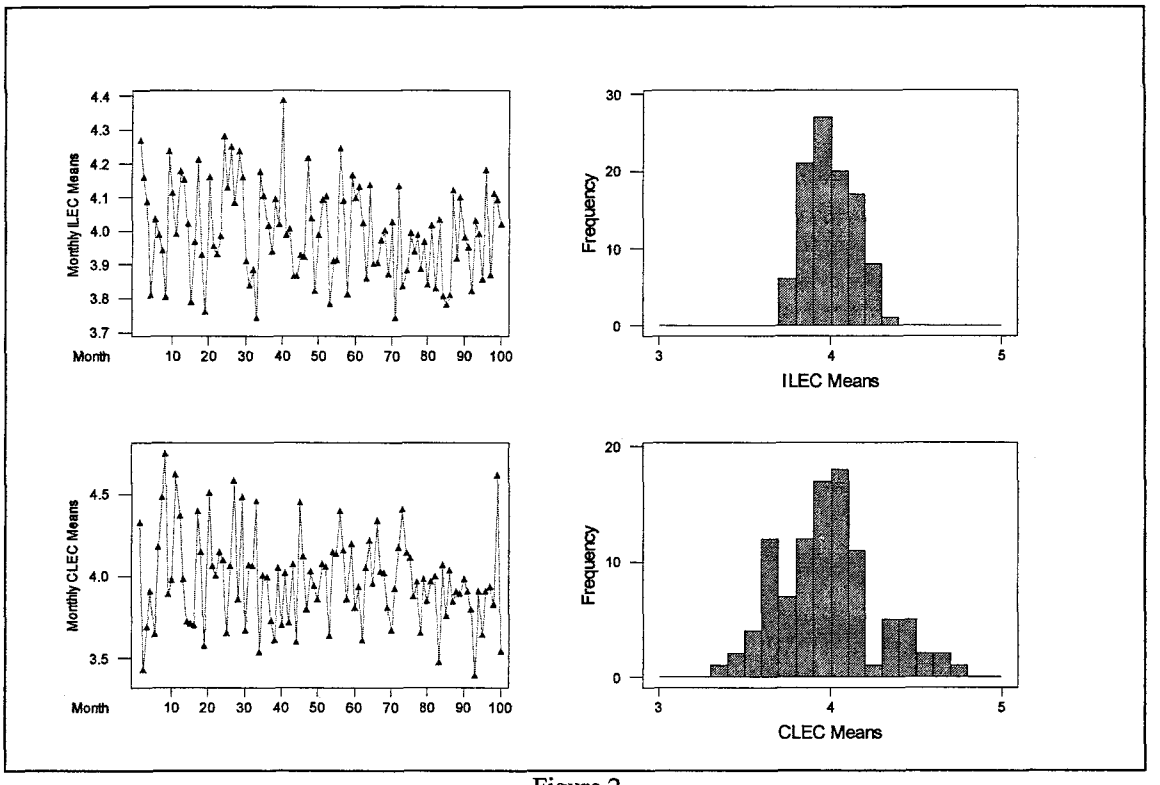


Figure 2.

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.*, $DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$) 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.[1]

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 (σ_{ILEC}).[2] 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".

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

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

---

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

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

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[3]; 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**
% 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**

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

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 ($\bar{x}_{ILEC}$ and $\bar{x}_{CLEC}$), and the sample standard deviations (σ_{ILEC} and σ_{CLEC}).

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

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

4. Compute the test statistic

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

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 $DIFF = p_{CLEC} - p_{ILEC}$, 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

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

4. Hence compute the test statistic

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

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 $r_{ILEC} = num_{ILEC}/denom_{ILEC}$ and $r_{CLEC} = num_{CLEC}/denom_{CLEC}$.

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

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

4. Compute the test statistic

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

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!

BEFORE THE
LOUISIANA PUBLIC SERVICE COMMISSION

IN Re: BellSouth Telecommunications, Inc.) Docket No. U-22252-C
Service Quality Measurements)

AT&T's Reply Comments

AT&T Communications of the South Central States, Inc. ("AT&T") submits the following Reply Comments pursuant to the Louisiana Public Service Commission's ("Commission's") procedural schedule outlining the dates for filing Comments and Reply Comments on Staff Initial Recommendation:

I. Statistical Methodology

Issue 1: What is the appropriate parameter delta for determining the difference in the means that should not be exceeded?

The parameter delta value should be set at or below 0.25. Otherwise, a large number of CLEC customers will receive an unacceptable quality of service without BellSouth being classified as out of compliance or being required to pay a remedy.

Consider a "counted" measure, e.g., Missed Installation Appointments. For each appointment a record is made as to whether that appointment was kept. Any customer whose appointment is missed will be dissatisfied and inclined to migrate to another company. Not all will do this, but some will. All customers with missed appointments are at risk. Therefore, a change of the Missed Installation Appointment rate from 5% to something higher will directly affect the rate at which customers are lost to a CLEC.

Due to unavoidable random variation, and the fact that each month there are only a finite number of observations, the result of the test that is applied may be in error. It may show violation when no systematic disparity exists, or fail to detect violation when it is present. If the critical value of the test is decreased, (i.e., the critical value is made more negative) the probability of the first kind of error is decreased, which BellSouth would like, and the probability of the second kind of error, which hurts the CLECs and their customers is increased. Moreover, the converse is also true. The interests of the two parties are directly opposed.

One way to resolve this difficulty would be to adopt some fixed value for the type 1 error probability, say 15% or 5% or 1%, and to let the type 2 error probability fall wherever it will, depending on how many observations there are each month. The effect of this approach would be that if there were only a few CLEC observations, the type 2 error probability would be large, while if there are many CLEC observations, the type 2 error probability will be small. Thus, all the variability (due to varying numbers of cases) is reflected in the value of the type 2 error probability, while the type 1 error probability is held fixed. This does not seem fair or reasonable for the CLECs.

The statisticians from AT&T and Ernst & Young jointly suggested, and both parties have agreed, that a more reasonable way to resolve this difficulty would be to agree on some specific degree of violation, and to arrange that the two error-probabilities are made equal for that degree of violation. This solution is called the "balancing" approach. It replaces the arbitrary choice of a level of type 1 error by a judgement of "material impact." The idea is that if for example, the BellSouth miss rate is 5% for its own retail operations, then there should be some proportion (x%)upon which the parties can agree that represents a material impact, while any proportion less than x% does not represent a material impact. Once this judgement has been made, and the observed sample sizes are known, the appropriate critical value for the test can be computed. The value thus derived will depend on the sample sizes. For example, assume there are many BellSouth appointments, and we have agreed to take x = 10%. In that circumstance, if there are 100 CLEC appointments, the critical value will be -0.96 and each of the two error-probabilities will be .168. If there are 300 CLEC appointments, the critical value is -1.67 and the error-probabilities are 0.048. And, for 1000 CLEC observations the critical value is -3.04 and the error probabilities are.0012. Thus, these critical values and error probabilities vary over a wide range, as the CLEC sample size changes. This is why there is concern about choosing any single value for the type 1 error probability.

The joint "Statisticians' Report" shows how a parameter called "delta" can be used to calibrate the degree of departure from parity. The formula expresses "delta" in terms of the two "Miss" proportions (BellSouth and CLEC) and appears in Appendix C of the Statisticians' Report.So all that is needed is for the two parties to agree on what value of delta constitutes "material difference." In the absence of such an agreement, this Commission must determine a reasonable value for delta – a value that will incent BellSouth to perform its obligations in a manner that will result in providing reasonable adequate service to end-user customers.

Clearly it is in BellSouth's interest to have a large value of delta (e.g., 1.00) because a large delta value makes the balancing critical values more extreme. Having a large value of delta also makes it less likely that any violation of parity will be detected. And as before, the CLEC's interest is directly opposed.

In comments to the FCC, AT&T presented a table illustrating what various values of delta mean, for "counted" measures such as Missed Appointments. (About half of all the VSEEM measures are of this type.) Specified below are two rows from that table:

Values of p(CLEC) for various values of p(BellSouth) and Delta.

	Delta		
p(BellSouth)	0.25	0.50	1.00
1%	5.0%	11.8%	31.9%
5%	11.8%	21.0%	44.0%

BellSouth recommends taking delta = 1.00. As illustrated by the table, BellSouth's judgement is that if BellSouth is missing 5% of its own customers' appointments, then the CLECs should not be allowed to claim anything less than 44.0% misses as a "material violation." The table also demonstrates that if BellSouth provides a 1% miss-rate for their own customers, then Delta = 1.00 implies the CLECs should have no basis for complaint until their miss-rate reaches 31.9%. The AT&T proposal is Delta = 0.25. With this value of delta and a BellSouth miss-rate of 5%, the CLECs could still receive more than twice as many missed appointments than BellSouth customers experience before a "material difference" is deemed to exist.

BellSouth has made the argument that delta should be taken equal to 1.00 because this makes the median critical value (for the sample sizes they consider) fall in a "reasonable" range, namely, near 1.645. This argument disregards the whole rationale for the "balancing" approach, the purpose of which is to make it possible to introduce "material impact" as the controlling factor.

It is important to note that BellSouth has not presented evidence, similar to the table provided above, showing the impact of any value of delta. BellSouth also has not provided similar evidence for measured variables. In its comments to the FCC, AT&T presented graphs showing the approximate impact of various values of delta on the Order Completion Interval measure. (An estimate of the shape of the distribution of Order Completion Interval was made because the published BellSouth reports do not give sufficient detail.) BellSouth has the data,

and presumably it has computed "impacts" such as those contained in the table. This Commission should order BellSouth to present the impacts.

The value of delta is of crucial importance, because under the Modified PIP and VSEEM III this value directly controls the point at which penalties will start to be assessed. Because the parties have not been able to agree on the value of delta, the Commission must decide a value that will result in the provision of reasonable and adequate service for the public. As demonstrated in the table above, BellSouth's currently proposed choice of 1.00 for the value of delta is not credible.[1]

II. Remedies/Enforcement Mechanism

Issue 2: Does the Louisiana Public Service Commission have authority to order implementation of a self-executing remedy plan without BellSouth's consent?

AT&T supports the Staff's conclusion that the Commission does have the legal authority to impose remedies.

Issue 3: Is BellSouth's proposed slope of ¼ in its VSEEM III proposal appropriate for determining the severity of the performance miss?

No. The Slope parameter enters into the VSEEM calculation as a factor that multiplies the "Parity Gap" to give the "Volume Proportion," except that if the result of this calculation is greater than 1, it is rounded down to 1. Thus if one has Slope = 1/4, any Parity Gap that is larger than 4 will lead to Volume Proportion = 1.

The VSEEM algorithm proceeds to multiply this "Volume Proportion" by the "Total Impacted CLEC Volume" for a counted measure such as Missed Installation Appointments. This so-called Total Impacted CLEC Volume is seemingly just the total number of missed CLEC appointments in cells for which the Z score is negative. The result is called "Affected Volume." This description is based on Exhibit D of the March 24, 2000 version of VSEEM. In an earlier

[1] AT&T is willing to calculate epsilon and psi associated with a delta of 0.25 to be used in parallel with 1.00 and its derivatives in the ensuing 9 months. The Statistician's Report showed how values of delta can be applied to proportions and rates. A fixed value of delta does not correspond to a fixed value of epsilon, since the relation between delta and epsilon involves the BellSouth proportion. Similarly, the relation between delta and psi involves the BellSouth rate. AT&T advocates using an agreed value of delta to determine the appropriate values of epsilon and psi, which are easily derived from the formulas in the Statistician's Report once the BellSouth proportions and rates are known.

version (February 2, 2000), the Volume Proportion was multiplied by the total number of CLEC cases (whether missed or not) in affected cells. In the May 16, 2000 BellSouth filing, some numerical inconsistencies in the MIA example were corrected. This is the latest version AT&T has seen. The Staff's Initial Recommendation refers to yet another version, in which a different critical value is used. The effect of these successive changes in the example is that the final "Affected Volume" has decreases from 133 (in February) to 29 (in March and April) to 15 in the version used in the Staff's Initial Recommendation). A question that can be asked at this point is which definition of "Affected Volume" was used in the remedy amount simulation that BellSouth performed during the workshops? If the simulations are indeed based on the February methods that give 133, then the subsequent changes may invalidate many conclusions reached by the staff.

The May 16, 2000 filing also gives an example of the calculation for a measured variable, Order Completion Interval. This example appears to be irrational . Many of the numbers are copied from the MIA table. The final "Affected Volume" is 29, as in the May MIA example, but it is completely unclear how this was obtained. In the first cell, the "Volume Proportion" has evidently been multiplied by 17 (the number of Missed Appointments in the May MIA table), but this number does not appear anywhere in the OCI table. Consequently, we do not understand the VSEEM proposal for measured variables.

A. The VSEEM calculation of "Affected Volume" is based on a flawed premise and therefore is incorrect. AT&T describes a much-simpler, alternative calculation that produces a correct "Affected Volume".

The VSEEM approach does not derive an appropriate "Affected Volume". A much simpler calculation will give the correct value. The discussion here uses the example presented by BellSouth and copied (partially) on pages 19-20 of the Staff's Initial Recommendation. AT&T will use only the columns headed:

nc (number of CLEC cases)

Ic (number of CLEC misses)

MIAi (BellSouth proportion of misses).

A calculation can be made of how many CLEC misses there would be in each cell, assuming parity is being provided. This calculation is made by multiplying the number of CLEC

cases times the BellSouth proportion. Hence, the excess number of CLEC misses can be derived. The numbers are as follows. The VSEEM "Affected Volume" results are given for comparison purposes.

Cell	nc	Ic	MIAi	Expected CLEC Misses	Excess CLEC Misses if Parity Holds	Positive Excesses Misses	VSEEM "Affected Volume"
1	150	17	0.091	13.65	3.35	3.35	4
2	75	8	0.176	13.2	-5.2		
3	10	4	0.128	1.28	2.72	2.72	1
4	50	17	0.158	7.9	9.1	9.1	4
5	15	2	0.245	3.68	-1.68		
6	200	26	0.156	31.2	-5.2		
7	30	7	0.166	4.98	2.02	2.02	2
8	20	3	0.106	2.12	0.88	0.88	1
9	40	9	0.193	7.72	1.28	1.28	2
10	10	3	0.16	1.6	1.4	1.4	1
					Totals	20.75	15

For example, in row number3, there are 10 CLEC cases and a BellSouth proportion of .128; so the expected number of CLEC misses is 10*0.128 = 1.28. Because the observed number of misses is 4, the excess number of misses is 4-1.28 = 2.72.

This calculation gives the proper "Affected Volume." It is simply the total excess number of CLEC misses, above that which would be expected if parity service had been given, ignoring cells where better-than-parity performance was observed. The "Affected Volume" is the number of CLEC cases that would need to be changed (from "Miss" to "Non-Miss") to bring all these cells into a parity configuration.[2]

[2] Although this calculation is done in the context of VSEEM, it should be noted that AT&T will later describe the correct method of doing a transaction based calculation. This method should take all

AT&T's method of determining "Affected Volume" is consistent with that which BellSouth has proposed for dealing with Benchmarked proportion measures (see Exhibit A of BellSouth's FCC Ex Parte, dated March 15, 2000.) BellSouth proposes to calculate the "Volume Proportion" simply as the difference between the benchmark proportion and the actual observed proportion, whenever this difference is positive. Multiplying this by the total number of CLEC cases will give the number of CLEC cases that would have to be changed (e.g., from Miss to Non-Miss) to bring the observed proportion up to the benchmark. The problem here is that Bell South gives itself a statistically based small sample size mitigation in VSEEM, which AT&T does not support.In contrast, the VSEEM calculation of "Affected Volume" is completely arbitrary, and not designed to measure a true affected volume.

B. The Linear Programming Calculation Performed By BellSouth To Justify Its Choice Of Slope = ¼ Does Not Approximate the True Affected Volume

The linear programming calculation does not give the number of transactions that should be remedied. This calculation asks, "What is the minimum number of CLEC customers that would have needed to receive improved service for the ILEC to have evaded detection?" This approach ignores the fact that improving CLEC-customer service just enough to avoid detection would still leave a disproportionately large number of CLEC customers with poor service. Any reasonable definition of Affected Volume would include that excess as well. The balancing critical value does not represent a parity situation.

The linear programming method used by BellSouth confuses the statistical principles of testing and estimation. The balancing critical value method was developed to limit the probability that the ILEC would need to make a remedy payment when they in fact are providing parity service. However, once a determination has been made that service is out of parity, the balancing critical value becomes irrelevant for estimating Affected Volume in a transaction based remedy calculation such as VSEEM.

Consider this analogy. Assume that the police who patrol a highway with a 55 MPH speed limit have the policy not to ticket anyone going 65 MPH or less. If an officer catches a motorist going 67 MPH, the fine is not based on going 2 MPH faster than you can get away with. The fine is based on going 12 MPH above the posted limit.

---

transactions that are outside of parity. There should be no statistical test. Even failed transactions that

C. The Linear Programming Calculation Does Not Justify a Slope of 1/4.

BellSouth has argued for Slope = 1/4 by presenting the results of their Linear Programming calculations. AT&T has not been allowed to see the full details of these calculations, and has difficulty relating the page of numerical results to the graph that is presented. For example, on the graph there are about five points with Volume Proportion very near to 0.5 and Parity Gap between 1 and 2; yet there are at most two such test cases in the table (rows 5 and 11).

Even if the linear programming version of "Affected Volume" were correct, the evidence presented would not justify a slope of 1/4. First, the line with slope 1/4 on the graph passes close to only a small fraction of the points; it is very far from most of the points.

Second, many of these CLEC counts are very small. One is able to deduce the numbers of CLEC cases in all the tests, by dividing the "Transactions paid" by the "LP Volume Proportion". Two are 1, two are 2, one is 3, five are 4. The maximum nCLEC is in test # 39, with n = 168; the corresponding point on the graph has the largest value of Parity Gap, and is very far from the line with slope 1/4. The next largest nCLEC is 91 (test # 32) and a corresponding point on the graph cannot be found.

Third, it is impossible that the line with a slope ¼ could fall near all the points, unless all the test cases had the same value of nCLEC. The reason is that the relation between the linear programming and VSEEM versions of Volume Proportion depends strongly on the number of CLEC cases. To see this, consider what would happen if the numbers of cases, and the numbers of misses, were to be multiplied by some factor, say 100. We would have 100 times as many BellSouth cases, 100 times as many CLEC cases, but the proportions of misses would be the same as in the actual data. The effect would be to multiply each cell Z by 10 (the square root of 100), and the aggregated Z would also be scaled up by this factor. The Critical Value would also be scaled up by a factor of 10, so the effect would be that the VSEEM Parity Gap would scale up by a factor of 10. The VSEEM "Affected Volume" would increase by a factor of 10*100 = 1000.

On the other hand, the linear programming calculation would scale up the number of Misses by a factor of only 100, and the linear programming version of "Volume Proportion"

---

occur in a measure that would have passed the test as a whole need to be remedied.

would not change at all. Thus the relation between the linear programming Volume Proportion and the VSEEM Parity Gap depends strongly on nCLEC.

D. The March Change to the "Affected Volume" of VSEEM III Warrants an Even Larger Slope

Even if a slope of 1/4 had been justified under the February version of VSEEM, subsequent changes introduce the need for a much larger slope. Beginning with the March version of VSEEM, the Affected Volume is computed by multiplying the volume proportion by the "Total Impacted CLEC Volume." Previously, the volume proportion had been multiplied by the "Total CLEC Volume." This change can dramatically reduce the results of the affected volume calculation. In the BellSouth example (page 1 of Exhibit D), the value dropped from 133 in February to 29 in March to 15 in the table shown in the Staff's Initial Recommendation --an 89 percent reduction. Without any change to the dollar multiplier from the Fee Schedule (the February and March exhibits and the Staff's Initial Recommendation all use the same multiplier of $100/unit), there is a corresponding reduction of 89 percent in the remedy payment. This is equivalent to applying a slope of 0.028 with the February method.
Therefore, even if a slope of 1/4 had been justified under the February version of VSEEM, it would no longer be justified unless per-unit amounts on the Fee Schedule are adjusted accordingly.From the above discussion it is clear that the VSEEM methodology for calculating remedies is heavily biased toward BellSouth and is therefore fatally flawed.

In response to comments pertaining to evaluating a Z score, the following comments apply:

On page 22 of the Staff's Initial Recommendation it is reported that an ALJ has found that "If the -1.645 threshold is crossed, the only thing which increases is the degree of certainty (above 95%) that a disparity occured. An increased Z score does not tell us anything about the size of the disparity or its importance to the parties."

This statement is not correct. For a measured variable, the Z-score is directly proportional to the degree of disparity, as measured by the parameter delta. In fact if we multiply Z by sqrt(1/m + 1/n) (where m and n are the two sample sizes, ILEC and CLEC), we get

an estimate of the value of delta that is operating. For counted variables it is a little more complicated, but when there are a large number of cases the relation holds here also.

Issue 4: Should a remedies plan require a continuous pay-out in relation to the escalation of the severity of the performance miss, and if so, how should that pay out be determined?

Yes, this is a desirable feature. Therefore, AT&T has incorporated this feature in its Performance Incentive Plan-Version 2.0. This is not a feature of VSEEM III. AT&T offers a reasonable and fair solution to developing a continuous payout function.

Once a 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. The AT&T proposal holds that the change in consequences amount should increase with each unit of increased severity. This behavior of consequences as a function of severity is most simply 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. In Table I, the quantity z^* is the (negative) balancing critical value for the submeasure, and the coefficients of the smooth consequence function are:

$a = 5625$

$b = -11250$

$c = 8125$

TABLE I

Range of modified z-statistic value (z)	**Performance Designation**	**Applicable Consequence ($)**
greater than or equal z^*	Compliant	0
less than z^* to $5z^*/3$	Basic Failure	$a(z/z^*)^2 + b(z/z^*) + c$
less than $5z^*/3$ to $3z^*$	Intermediate Failure	
less than $3z^*$	Severe Failure	25,000

Note that the concept of Intermediate Failure is retained even though the formula used to calculate it and the Basic Failure is the same. This enables classification of failures in a more refined manner if it is desired to use these classifications for more general performance evaluation purposes such as compliance determination.

Issue 5: What is the appropriate level of disaggregation to include in a remedy plan and for purposes of determining the amount of remedy payments?

AT&T supports the Staff's recommendation for the following additional levels of disaggregation to address "like-to-like" comparisons:

1. Additional Product Disaggregation

   Resale ADSL

   UNE xDSL

   UNE Line Sharing

2. Disaggregation of partially and manual LSRs for the Reject Interval & FOC Timeliness measures.
3. Disaggregation of flow-through measure by residence, business, UNE & LNP.

Although the staff recommendation represents an improvement of BellSouth's proposed disaggregation, the level of disaggregation required to ensure "like-to-like" comparison is not reflected in this initial recommendation. Disaggregation should be by interface type. If TAG data is lumped together with LENS data, the performance of the TAG interface will be obscured. Pre-order query type disaggregation is important because a request for something simple like a phone number may require less response time than a request for something more complex like a due date reservation or loop makeup information. The basic principle of product disaggregation is that each product should be tracked separately.

As an example, DS1 and DS3 have differing provisioning and repair intervals and complexities that require separate reporting. Separating BRI ISDN and PRI ISDN is appropriate for the same reason. UNE loop types such as analog voice-grade loops, digital loops, and UCL should be disaggregated because BellSouth's performance will vary for each type of loop. Example of service order activities include new service installations and service migrations without changes. Because these different service order activities involve different processes, they should be reported separately. Different volume categories can result in different intervals.

Lumping together different kinds of troubles leads to meaningless results. For example, data for the mean time to restore service for a trouble requiring dispatch to the customer's premises should not be included in the same data set as the mean time to restore service for a trouble not requiring dispatch.

Aggregating trunks designed at different blocking thresholds could hide serious blocking problems by averaging trunks designed to block at 2%, 1%, or .5%. Different types of collocation take different amounts of time to provision. For example, provisioning a cageless collocation space should require substantially less time than provisioning a caged collocation space.

Issue 6: Is it appropriate to determine volume proportions at the aggregated level and then determine pay out at the disaggregated level as proposed in BellSouth's VSEEM III remedy plan?

No. BellSouth's method of determining "affected volume" is flawed. See AT&T's response to Issue 3.

Issue 7: Should any remedy plan adopted by the Louisiana Public Service Commission contain procedural caps, absolute caps or no caps at all?

AT&T supports Staff's recommendation of a procedural cap.

Issue 8: Should remedy payments apply for late, incomplete, or erroneous performance reports and raw data?

AT&T supports Staff's recommendation for general remedies although the remedy amount seem too low.

Issue 9: Should remedies apply to all measures; and if not, to what performance measures should remedy payments apply?

Staff's recommendation to add Average Completion Notice Interval is very significant. However, if metrics designated as "parity by design" cannot be confirmed, then these measures should be added to the plan. Note that if a parity by design metric subsequently shows itself to be biased, then the Commission should provide for strong consequences for Bell South.

Importantly, Staff's recommended 80 metric set is smaller than New York which has 120 and much smaller than Texas which has about 1800 submeasures.

Issue 10: How should Tier II (CLEC aggregate misses) remedies be applied?

AT&T supports the Staff's recommendation to not allow the slate to be wiped cleaned at the end of a calendar quarter as a form of foregiveness. Staff identified 3 forms of foregiveness in BST's VSEEM III which are as follows:

- Development of truncated Z score, producing a less negative Z score than if positive performance was given no weight
- A (¼) slope parameter used to calculate its volume proportion
- The delta value defining the alternative hypothesis

No forgivenesses should be afforded in a balancing type plan. AT&T has enhanced its original Tier II calculation which is described in detail in AT&T PIP Version 2.0. In the enhanced version of PIP AT&T accomplishes, in a fair manner, two important goals recommended by staff during workshop discussions: 1) AT&T incorporates the Balancing Methodology for all parity submeasures, and 2) AT&T gives a continuous schedule for the escalation of consequences with severity of performance degradation for both parity and benchmark submeasures.

Issue 11: Should there be a Tier III penalty suspending BellSouth's ability to market long distance services; and if so, how and under whatcircumstances should a Tier III penalty be invoked?

In the event that VSEEM III's Tier III is never triggered, but BellSouth, nevertheless, is providing severely discriminatory support, AT&T recommends that the Commission should hold an expedited hearing to determine if it should recommend to the FCC that BellSouth's approval to provide inter-LATA long distance service be revoked.

Issue 12: Should remedies accrue on a "per transaction" or "per measure" basis?

Remedies should accrue on a per measure basis. However, if the Commission wishes to use a per transaction method as opposed to a per measure method, then any and all statistical tests are precluded because each failed transaction will therefore require a remedy. This remedy will accrue regardless of whether the transaction comes from a submeasure that would have passed a statistical test or not. What this means is that all transactions in violation should incur a remedy.

The basic tenet of the BellSouth VSEEM payment plan is that payments should depend on the number of affected cases. If this philosophy is accepted, AT&T suggests the following simplified version of VSEEM:

The basic principle is that for each measure, the payment should depend on the extent to which CLEC customers have been adversely affected, beyond what parity treatment would give. This can be determined directly from the data. For a counted variable, for example Missed Installation Appointments, one would count the number of CLEC cases that are in excess of what would be expected if the CLEC proportion was equal to the BellSouth proportion. Call this number the "Excess". For example, if in one cell BellSouth missed 8% of its own appointments, and 23 out of 100 CLEC appointments, the Excess would be 23 - (8% of 100) = 15. This calculation is done separately for each cell. If in some cell the CLEC received better-than-parity service, the Excess would be negative, and would be ignored as usual. The total Excess for this measure is obtained by adding (over cells) the positive Excess values.

The same approach can be applied to benchmarked measures. For example, for the measure "Percent response received within X seconds", suppose the benchmark is 90%. That is, one would allow 10% of the cases to be failures. If the data shows that 3 out of 12 CLEC cases are failures, we count the excess number of missed cases as 3 - (10% of 12) - 3-1.2 = 1.8. We do not round off to an integer value (except possibly at the end, after aggregating). Once the total "Excess" for a measure is determined, it is translated into a dollar figure, according to a table similar to that provided by staff, to get the payment amount. For counts and benchmarks, this approach conforms to the VSEEM idea that one should "count cases", but does it in a much more direct way. There is no need for any Z-scores, or modified or truncated Z's. There is no testing involved, so the whole discussion regarding balancing critical values and the value of delta becomes irrelevant. Also, it avoids the necessity of dealing with aggregation once agreement has been reached concerning the lowest level of disaggregation for a cell. To aggregate one would simply add up the "Excesses".

In order to get a dollar figure from the total Excess number, one could have a simple linear rule, i.e., $X per case. This is may not be satisfactory given that the payment should escalate faster than a linear function (larger violations are more serious. When violations become very numerous, the whole CLEC operation is in peril.) There is room for argument and negotiation as to just what function of the total Excess should be used. To start with, a quadratic function is suggested. Guidance as to what makes sense can be obtained by running

some simulations, in which one would have various scenarios (patterns and degrees of violations) and see what the expected payment would be for each scenario. Since the rule is so simple, this should not be difficult. To get useful results, one would need to have good guesses as to what the CLEC sample sizes and BST percentages might be, for each measure.

For a measured variable such as Order Completion Interval, there are two ways to proceed. First, one could transform the measure to a counted variable by recording only whether the OCI was greater or less than some threshold value. This approach should be rejected, because once an order has been delayed past the threshold, BellSouth would have no incentive to complete the order. BellSouth's lack of incentive results from the lack of any additional penalty for further delay. A better approach is for each CLEC case with a measurement that is above the BellSouth average, to determine the excess measurement, and to add the positive excesses over all CLEC cases. For OCI, this would give the total "Days Delayed". As before, this has to be translated into a dollar amount.

Issue 13: Should a remedy plan include a CLEC market penetration adjustment, and if so, how should such an adjustment be implemented?

Staff has recommended a market penetration methdology. However, this methodology may not generate remedies significant enough to address the Staff's intended purpose. The VSEEM III calculation already minimizes the remedy amount. Therefore, tripling an already inadequate amount is not going to result in the motivation that is required to influence BellSouth to provide nondiscriminatory behavior. Given the inadequacies of the VSEEM III calculations, it would seem to be inappropriate for the market penetration adjustment methodology to be dependent on VSEEM III's current Tier II calculation. If Staff in insistent on its recommendation, then Staff should order BellSouth to fixed its remedy calculation such that the remedies are meaningful. AT&T has proposed a market penetration adjustment in the PIP which generates the level of remedies to motivate proper behavior. If implemented CLEC business plans will evolve to include BST states because CLECs will see that there is a chance for doing business. It will firm the commitment of Bell South to opening markets.

Issue 14: Should remedies apply to performance measures that have been shown through an independent third party audit to qualify as "parity-by-design"?

See AT&T's March 20, 2000 and April 20, 2000 responses filed in this docket.

Issue 15: Should remedies apply to performance measures included in BellSouth's Service Quality Measurements document but which have not been specifically requested or endorsed by the CLECs?

See AT&T's March 20,2000 and April 20, 2000 responses filed in this docket.

Issue 16: Should remedies apply to performance measures that are shown to be duplicative of or "correlated" with other measures?

See AT&T's March 20, 2000 and April 20, 2000 responses filed in this docket.

Issue 17: What measures in the BellSouth's SQM document are duplicative or correlated with other measures?

BellSouth has provided a preliminary "Statement Concerning Correlated Measures" in which several correlations are calculated. Unfortunately these calculations are flawed; they do not provide valid estimates of the correlations between the measures. The problem is the same problem that affected the original BellSouth method; cells that have fixed structure are treated as though they were random. To illustrate this difficulty, assume there are two measures that are exactly uncorrelated. Also assume that in all but one cell, there is exact parity. Then for these cells, the Z-values for each measure will be close to zero (+-1). Now assume that in the final cell there is gross violation of parity, to the same extent in both measures. Then the BellSouth calculation will give a large positive value for the "correlation" between the measures. BellSouth's calculation treats differences between cells, which may be due to a systematic pattern of violation of parity, as though they were due to "correlations" between the measures. However, without seeing the data, one cannot judge whether a valid estimate of correlation can be made from a single month's data.

Issue 18: Should remedies apply to ordering and LNP performance measures that reflect manual and partially mechanized processing?

AT&T supports the Staff's inclusion of partially mechanized and manual LSR's for ordering and LNP performance measures.

Issue 19: What is the appropriate adjustment for small sample sizes when a benchmark is used to determine the standard of performance?

AT&T has proposed a small sample size adjustment in the original PIP. This adjustment is retained in the modified PIP. BellSouth then produced its own "me too" adjustment table based on statistical calculation after AT&T gave its simple rounding up by one point and compare method. BellSouth's statistically based method is inconsistent with a Benchmark limit concept where all statistical variation has already been included in the value of the benchmark

proportion. Using additional statistical methods amounts to a double mitigation. This is explained in the section of PIP that proposes the small sample size adjustment table.

Issue 20: When should any remedy plan adopted by the Commission go into effect?

The remedy plan should go into effect as soon as possible, well before 271 approval. In fact the remedy plan and its designations and definitions are a reasonable way to establish compliance before approval as well as helping to prevent backsliding afterwards.

Issue 21: Should BST be required to include a "root cause analysis" as provided for in Sprint's penalty comments?

If the Staff decides to implement such an analysis, it should be made clear that the analysis is not substituted for the remedy.

Issue 22: What are the appropriate performance standards to be applied in the remedy plan?

The standards in general are parity and benchmark. AT&T recommends parity when there is a retail analog and benchmark when not. The benchmark should have a level and proportion which give an efficient provider a meaningful opportunity to compete.

III. Measures, Disaggregation and Business Rules

Issue 23: What is the appropriate level of disaggregation to include in the BellSouth SQM document?

AT&T Position: The Staff Recommendation addresses the issue of product disaggregation in Issue 53. See AT&T response to Issue 53.

Issue 24: Should there be a measure or measures for BellSouth's performance on "hot cuts", and if so, what are the appropriate measures?

AT&T supports the Staff-recommendation for the addition of early and late cuts to BellSouth's hot cut measures and requests that these critical measures also be included in the remedies plan. AT&T encourages the Staff to adopt the remainder of the hot cuts requested by AT&T, or minimally those in place in Texas and New York. For example, the percent of provisioning outages was cited in both the New York and Texas 271 FCC orders as a critical measure, but BellSouth has no such measure. AT&T also notes that BellSouth also proposed a

metric in Georgia to measure the percent of hot cut provisioning troubles occurring within 7 days, and recommends that the Staff recommend that this measure be added in Louisiana.

Issue 27: Is it appropriate to exclude permit time from the calculation of the average time to provision a collocation arrangement?

AT&T supports the Staff's recommendation to exclude permit time from the application process, but to allow for a process for BellSouth to request a waiver.

Issue 28: What is the appropriate number of days for reporting and measuring provisioning troubles?

AT&T requests that the Staff reconsider its decision to adopt BellSouth's recommendation of five days. In addition to its previous comments, AT&T offers the following:

On June 27, 2000, BellSouth filed its proposed service quality measures in Georgia which includes a measure of provisioning troubles within 30 days. BellSouth also proposed a metric in Georgia to measure the percent of hot cut provisioning troubles occurring within 7 days. AT&T is supportive of BellSouth's timeframes proposed in Georgia and requests that they be implemented in Louisiana as well. Due to BellSouth's inability to provide a study in support of its claims that the majority of the troubles occur in the first 4 days, AT&T requests that BellSouth continue to report its performance for the 30 day period until it can produce a new study in support of its claims.

Issue 29: Should time be included in the formula for calculating the Percent Missed Installation Appointments measurement?

See AT&T comments in Issue 24. If hot cuts are the only time-specific appointments BellSouth offers CLECs and those are adequately measured, and it offers none to its retail customers, then AT&T supports the Staff Recommendation not to require a time for this measure.

Issue 30: When should the measured interval begin and end for pre-order response time measurement?

AT&T does not agree with the Staff that "the issue here boils down to whether or not the time required to pass though BellSouth's security safeguards still gives CLECs a meaningful opportunity to complete." While an opportunity to compete is certainly relevant and critical, the

appropriate standard to be applied for this analogous service is non-discrimination. As AT&T pointed out in its reply comments, BellSouth's claim about security processes applicable to CLECs are inappropriate as the query either does not seek proprietary data or it applies both to BellSouth and CLECs. Ironically, BellSouth argues for a limited measure of this interval due to security safeguards at the "BellSouth Gateway", and then **ALSO** argues for parity + 4 seconds (See Issue 51) to accommodate the security issues it just eliminated with its limited measure (legacy to application instead of query to response). AT&T strongly believes that this measure is structured inappropriately[3], but is willing to support the Staff Recommendation that the Commission use the data from the KPMG audit in Florida as input into the potential revision of this measure.

Issue 31: Should BellSouth be required to post its own scheduled hours of OSS availability on its web-site as it currently does for CLEC OSS availability?

AT&T still believes this to be an open issue. The fact that this measure has a benchmark does not alleviate concerns regarding this measure because it does not evaluate absolute availability, but availability relative to a schedule controlled by BellSouth. Unless CLECs have information regarding BellSouth's own schedule, there is no method of evaluating the reasonableness of BellSouth's OSS availability schedule for CLECs.

Issue 32: Should error-free local service requests that fall out for manual processing be excluded from the percent flow-through measurement?

AT&T supports the Staff's recommendation that BellSouth be required to enhance its flow-through offerings. AT&T also supports the Staff recommendation that BellSouth be required to implement a flow-through metric that includes its LSRs designed to fall out for manual handling. AT&T notes that the Staff believes that the issue of whether or not the CLECs are treated the same as BellSouth retail customers will be handled as part of the KPMG audit of BellSouth's SQM. However, the Staff made no recommendation regarding the findings of the audit. AT&T requests that the Staff also recommend that BellSouth be required to remove its exclusion of any service from the flow-through measure for which KPMG finds that BellSouth's own retail services can be ordered without fall-out for manual handling subsequent to order submission.

[3] See Testimony of Jay Bradbury filed June 27, 2000 in Georgia docket 7892-U.

Issue 33: Should BellSouth service representatives be permitted to assign errors as belonging to a CLEC or BellSouth for purposes of the percent flow-through measurement?

AT&T supports the Staff Recommendation that an independent work group should be responsible for classifying flow-through errors. AT&T requests that the Staff also recommend that BellSouth implement this change no later than sixty days following a Commission Order.

Issue 34: When should the measured interval end for the rejection interval measurement?

AT&T does not support the Staff's Recommendation that the end time should be when the reject is queued to be returned to the CLEC. AT&T is aware of no evidence in the record to support the conclusion that "there should be no significant or material time difference between BellSouth's method of determining when the interval ends and the CLECs' proposal for making this determination." Further, for non-mechanized orders, AT&T never requested that BellSouth manually monitor fax machines. It is AT&T's understanding that BellSouth uses a fax server, not a fax machine, to transmit its rejects. AT&T is simply requesting that that the time stamp from the fax server be used as it is available data that more accurately captures the reject interval than the current process of using LON.

Issue 35: Should all versions of LSRs be used in the denominator of the rejection interval and percent rejected service requests measurements?

In addition to the Staff Recommendation that the parties be notified when all versions of rejections are included in the denominator of this measure, AT&T requests that the Staff should also recommend that BellSouth be ordered to correct this deficiency no later than 60 days of a final order. Further, BellSouth should be required to update its SQM to reflect that this data is not included until the deficiency is corrected.

Issue 36: When should the measured interval end for the FOC interval measurement?

See AT&T response in Issue 34.

Issue 37: What time should be excluded from the calculation of the FOC and rejection interval measurements?

AT&T supports the Staff's proposed changes to the exclusions for this measure, except for the decision to exclude hours outside normal operating hours for partially mechanized LSRs.

AT&T notes that in Issue 18, the Staff concluded that "for the majority of the partially mechanized orders, only BellSouth can control the movement toward mechanization." AT&T agrees with the Staff and believes that BellSouth should therefore be accountable for the hours it allows a partially mechanized order to remain in an unprocessed state. AT&T further notes that in the SQM proposal filed by BellSouth in Georgia on June 27, 2000, BellSouth did not include partially mechanized LSRs in it is exclusions for the FOC and rejection interval measures. AT&T recommends that Louisiana treat its LSRs the same as BellSouth recommended in Georgia.

Issue 38: Should orders cancelled after the due date be excluded from the held order interval measurement?

AT&T notes that in the SQM proposal filed by BellSouth in Georgia on June 27, 2000, BellSouth did not exclude cancelled orders from the held order interval measure. AT&T recommends that Louisiana adopt the exclusions for this measure that BellSouth recommended in Georgia. Additionally, AT&T seeks clarity on the weight given the FCC's NPRM in the Staff's analysis of these issues. In Staff's response on this issue, it was noted that BellSouth's position was consistent with the FCC's NPRM. However, in Issue 36, when AT&T quoted the NPRM as consistent with its recommendation, the Staff responded "while the FCC proposed this definition of the FOC interval, no order has resulted from that NPRM, and this definition has not yet been adopted." (Also see Issue 44) The Staff must be consistent regarding its reliance (or lack of reliance) on the NPRM.

Issue 40: Should orders submitted to BellSouth via non-mechanized methods be included in the jeopardy order interval and percent orders given jeopardy notice measurements?

AT&T supports the Staff's recommendation that jeopardy notice measurements should include all services requests, whether mechanized, partly mechanized, or non-mechanized.

Issue 41: When should the measured interval begin and end for average completion interval measurement?

AT&T does not support the Staff's recommendation and requests that it reconsider its recommendation based on the following information. The Staff recommended that BellSouth's definition be accepted, as it allows CLECs and BellSouth both to examine the provisioning time as a separate interval. However, the CLECs are not seeking a measure of how long it takes to

provision its orders, they are seeking how long is takes to complete its orders, as do the equivalent measures in Texas and New York[4]. Allowing the measure to remain structured as BellSouth recommends also prevents any meaningful retail analog from being used, as BellSouth provides no FOC interval to include. Therefore, BellSouth can issue its order up to two days or more (depending on the service) after a CLEC order, give it the same due date as a CLEC order, and still make a claim of parity of completion interval when no such parity exists for the CLEC's customers. AT&T requests that minimally the Staff recommend the start time for this measure be modified to begin with the receipt of a valid LSR in order to allow parity determinations to be made regarding the interval provided to a CLEC customer and a BellSouth customer.

Issue 42: Should orders submitted to BellSouth via non-mechanized methods be included in the average completion notice interval measurement?

AT&T does not support the Staff's position that BellSouth should not be required to measure its performance for the provision of completion notification on non-mechanized orders, and requests that it reconsider its recommendation. In its analysis, the Staff notes that it concurred with CLECs in Issue 40, and also notes that CLECs used the same rationale for this issue. The Staff goes on to say because BellSouth has provided two means of receiving notice, BellSouth has provided the CLECs with sufficient means to ascertain that the installation has been completed. AT&T is perplexed by this rationale however, as this is not a completeness measure, i.e., how often do CLECs get completion notices, it is a timeliness measure, e.g. what is the interval between completion of work and notice of completion. Additionally, not only do CLECs need completion information, the Commission needs it as well to monitor BellSouth's performance and in the implementation of its remedies plan.[5] AT&T requests that the Staff require that BellSouth not only provide completion information for non-mechanized orders, but that it measure and report its performance as well.

Issue 43: When should the measured interval end for the average completion notice interval measurement?

To date, BellSouth has not provided the end time for this measure. To state that the "end time is the time stamp the notice was submitted to the CLEC/BST system" provides no useful

[4] The Staff noted that the BA-NY measure is the equivalent of BA's TSOCT. However, BellSouth's TSOCT has no performance standard and is diagnostic only.

information. In other measures, BellSouth states which system, e.g. LEO, SOCs, LON, LENS, etc. AT&T requests that BellSouth provide the name of the system from which it extracts its time stamp to stop the timing for this measure so that AT&T can determine if it agrees with this measure.

Issue 45: Should trouble reports greater than 10 days be excluded from the maintenance average duration measurement?

AT&T supports the Staff Recommendation that these troubles not be excluded.

Issue 46: Should the following LNP measures: percent rejected service requests, the rejection interval, the FOC interval, percent missed installation appointments and average disconnect timeliness, and total service order cycle time, exclude non-mechanized orders?

AT&T supports the Staff Recommendation that non-mechanical orders be included in the five LNP measures at the LNP (without loop) level. AT&T seeks clarity on Staff Recommendation for the six measures, Average Disconnect Timeliness Interval. The Staff recommends that "all LNP measures should be reflected in it." Does the Staff intend that all disconnects associated with LNP, ordered electronically and manually, for both LNP stand-alone and LNP with loop be included? It is of obvious concern to CLECs that timely disconnects be processed for all those order scenarios, and AT&T recommends that the Staff Recommendation make clear that is its intention.

Issue 47: Should the following LNP measures, percent missed installation appointments and average disconnect timeliness, and total service order cycle time, exclude LENs orders?

AT&T supports the Staff Recommendation. Also see AT&T comments in Issue 46.

Issue 49: Should the Invoice Accuracy measure be determined by adjustment dollars?

If BellSouth is allowed to continue this method of measuring, as the Staff recommends, AT&T requests that BellSouth also be required to submit a report of the dollar amount of claims submitted and the dollar amount of adjustments granted, and the number of claims submitted and the number of claims denied. This will give the CLECs and the Commission information with which to monitor BellSouth's rejection of part or all of CLEC claims.

[5] AT&T notes that the Staff recommends the addition of the Average Completion Notice Interval to the

IV. Analogs and Benchmarks

Issue 51: What are the appropriate analogs or benchmarks for the following Pre-Ordering measures: (a) Average OSS Response Time, (b) OSS Response Interval, and (c) OSS Interface Availability?

OSS-1 Average Response Time and Response Interval (Pre-Ordering)

AT&T disagrees with the Staff Recommendation. BellSouth's eleventh hour reversal of its position should not be accommodated, as it provided no evidence to support its position. Further, its unsubstantiated request for parity plus x was not accommodated in the Georgia and Florida interim measures established for the third party tests. It is more appropriate to change the analog after the study, if appropriate, than to change it before the study results are available, as the Staff has recommended. Interestingly, BellSouth does not point to its previously reported results to support its request. In reviewing its first retail data reported in months (May data), it becomes apparent why is does not.

SYSTEM	RNS	LENS	TAG
RSAG	1.06 seconds 1.56 seconds	1.07 seconds 1.31 seconds	1.32 seconds .93 seconds
ATLAS	.91 seconds	.75 seconds	1.50 seconds .78 seconds 2.34 seconds
DSAP	1.01 seconds	.43 seconds	.60 seconds
CSR	2.97 seconds	3.34 seconds	.95 seconds 8.02 seconds
Service/Features	.55 seconds .20 seconds .27 seconds .46 seconds .1.00 seconds	.60 seconds 2.17 seconds	Not reported

BellSouth provided no data justifying its new analog plus plan. Its performance data does not justify it. The Staff Recommendation should be revised to make the standard for this measure a retail analog.

---

measures to which penalties would apply.

OSS-2 Interface Availability (Pre-Ordering)

AT&T supports the Staff Recommendation of 99.5%.

Issue 52: What are the appropriate analogs or benchmarks for the following Ordering measures: (a) Percent Flow-Through Service Requests, (b) Firm Order Confirmation, (c) Reject Interval, and (d) Speed of Answer in Ordering Center?

O-1 Percent Flow-Through Service Requests

Although AT&T strongly believes that the Staff has established too lenient[6] benchmarks for this measure, AT&T strongly supports the Staff recommendation that improvement to the benchmarks be implemented.

O-6 Firm Order Confirmation Timeliness

AT&T does not support the Staff's recommendation because it rewards BellSouth's previous discriminatory behavior and institutionalizes the unnecessary designed delay BellSouth has created for CLEC orders. A study recently conducted by AT&T reveals that AT&T's partially mechanized orders languish untouched in BellSouth's systems for more than a day in the majority of cases, only to be processed in 20 seconds to 2 hours once "claimed" by a BellSouth employee. (See Testimony of Jay Bradbury filed June 27, 2000 in Georgia docket 7892-U). Further BellSouth's threshold of 85% allows for gaming of the system[7]. AT&T is unaware of any other RBOC being allowed to perform at such low levels. AT&T disagrees that performance standards such be established based on what BellSouth is now achieving. That approach does not appear to consider any standard of non-discrimination or a meaningful opportunity to compete. AT&T urges the Louisiana Commission to require that BellSouth perform at a level of FOC performance that is required of other RBOCs. Although the Staff's

[6] The Staff's recommendation establishes that BellSouth's ordering OSS only have to work as designed 80% of the time for business resale and UNEs.

[7] AT&T objects to the low thresholds such as 80% and 85% that BellSouth recommends in its benchmarks, which allow BellSouth to provide poor service to CLECs. Thresholds or "proportions" such as 95% are established to allow for some variation that may occur in processes, not to render the required performance (such as 24 hours) meaningless by allowing excessive exceptions. The ranges proposed by BellSouth are indicative of one of two things, either an unstable process, or an attempt to game the system by agreeing to a level of required performance that implies one level of service such as 24 hours, but avoiding having to reliably perform at that level by establishing a low threshold such as 85.

recommendation requires some level of improvement of BellSouth over time, the required improvement is not adequate to meet CLEC needs.

O-5 Reject Interval

See AT&T Comments regarding FOC Interval.

Issue 53: What are the appropriate analogs or benchmarks for the following Provisioning measures: (a) Average Completion Interval, (b) Order Completion Interval Distribution, (c) Mean Held Order Interval, (d) Held Order Distribution Interval, (e) Average Jeopardy Notice Interval, (f) % of Orders Given Jeopardy Notices, (g) Percent Missed Installation Appointments, (h) Percent Provisioning Troubles within 4 days of Service Order Activity, (i) Total Service Order Cycle Time, and (j) Average Coordinated Customer Conversion Interval.

Level of disaggregation: In its recommendation, the Staff rejected AT&T's recommended level of disaggregation, stating that it was "excessive" and that the level of product disaggregation proposed by BellSouth "with minor additions"[8] will be adequate. AT&T notes that BellSouth apparently now believes that the recommendation it made to the Louisiana Commission is inadequate, at least inadequate for Georgia. Following are categories of UNEs offered in Georgia but not offered in Lousiana:

Analog loop
UNE Digital Loop <DS1
UNE Digital Loop > DS1
A breakout of combos between loop and port and other combos
UNE-ISDN – UL

Retail Analogs: The Staff appropriately indicates in its recommendation that the proper level of product disaggregation must be determined before one can determine the (associated) proper standards. However, the Staff then appears to focus solely on the product disaggregation that should occur on the CLEC side, with concerns about disaggregating to a level that would result in categories populated with little or no data. No consideration appears to be given the types of data that will populate the retail side of the product disaggregation, which then becomes the retail analog. For example, the Staff supports BellSouth's simple and deceptive approach of comparing three types of UNE design to BellSouth's overall retail design. However, retail design compares the many products in BellSouth's mature non-POTS market with the relatively

[8] The Staff recommended the addition of resale ADSL, xDSL loops, and line sharing.

few types of UNEs currently being ordered in the emerging CLEC market. Predictably, the results are not comparable. For example, a review BellSouth's May SQM results shows that UNE Design-dispatch order completion interval was 13.02 days, while the retail design dispatch was 25.56 days, nearly twice as long. A review of prior months and of the non-dispatch category reveal similarly skewed results. These are clearly not analogous processes and should not be used as a retail analog. AT&T strongly recommends that the Staff remove the problematic designation of design from the product disaggregation and the retail analogs. (Also see AT&T recommendation below).

Dispatch Orders: The Staff Recommendation concurs with BellSouth's position that UNEs should compared with dispatch only, citing the need for physical work for most UNE provisioning. This approach, however, is inadequate for the following reasons. Much of the work for UNE provisioning is "dispatch in" at the central office where a technician can travel once and provision multiple orders. Dispatch out, on the other hand, involves more complicated and time consuming activities such as visiting the end-users premises. It also appears that BellSouth is even recommending dispatch for loop and port combos, even though in many cases this product requires no physical work. For example, AT&T believes that much of its loop/port activity will involve customer migrations from BellSouth to AT&T, feature or other software changes, as well as requests for new services that will not require a dispatch. To compare these types of orders, which require no dispatch to retail residence and business dispatch is totally inappropriate. While BellSouth does not currently report loop/port combo data, the following information from its May SQM is nonetheless instructive. The retail residence dispatch order completion interval was 5.74 days, the non-dispatch was .93 days. The retail business dispatch order completion interval was 7.93 days, the non-dispatch was 1.58 days. Clearly, the non-dispatch orders for loop/port combos should not be compared to retail dispatch.

To ensure like processes are being compared, AT&T requests that the Staff recommend that BellSouth disaggregate as follows for the "dispatch" category; dispatch in, dispatch out, and no dispatch for both CLECs and BellSouth.

Finally, AT&T reiterates that its recommendation detailed in its earlier comments regarding disaggregation and standards be adopted, but if the Staff does not concur, AT&T

recommends at minimum alternative in Attachment 1 as a starting point for disaggregation and standards for provisioning and repair measures.

Issue 54: What are the appropriate analogs or benchmarks for the following Maintenance & Repair measures: (a) Customer Trouble Report Rate, (b) Percent Missed Repair Appointments, (c) Maintenance Average Duration, (d) Percent Repeat Troubles within 30 Days, (e) Out of Service > 24 Hours, and (f) Average Answer Time in Repair Center?

See AT&T response in Issue 53.

Issue 55: What are the appropriate analogs or benchmarks for the following LNP Measures: (a) Average Disconnect Timeliness Interval, (b) Percent Missed Installation Appointments, (c) Firm Order Confirmation, (d) Percent Reject Service Request, (e) Average Reject Interval, (f) Total Service Order Cycle Time, and (g) Percent Flow-Through?

LNP - Average Disconnect Timeliness Interval-

AT&T supports the Staff's recommendation and notes that this measure evaluates a key customer-affecting activity for BellSouth. For example, lack of timely disconnects in those instances in which a ten digit trigger is not applied will result in a customer not being able to receive intra-switch calls.

LNP - Percent Missed Installation Appointments

In Issue 46, the Staff recommended that non-mechanized orders be included in this measurement. However, in this Issue, the Staff appears to support that BellSouth report only "LNP electronic-LSR orders". AT&T requests that the Staff clarify that this measure does include stand-alone LNP non-mechanized orders. If not, it does not appear that this type of order activity will not be reported or measured by BellSouth.

LNP- Firm Order Confirmation Interval

AT&T supports the Staff recommendation for a benchmark of one hour for (fully) mechanized orders. See AT&T comments in Issue 52 regarding partially and non-mechanized FOCs.

O-1 Percent Flow-Through Service Requests

See AT&T comments in Issue 52.

Issue 56: What are the appropriate analogs or benchmarks for the following Collocation Measures: (a) Percent Missed Due Dates, (b) Average Response Time, and (c) Average Arrangement Time?

C-1 Average Response Time

The Staff acknowledges that BellSouth's offered intervals are liberal when compared with other states, but also indicates that BellSouth appears to have difficulty in meeting these intervals. Therefore, the Staff recommends that CLEC requests for better performance should be rejected and BellSouth's proposal adopted. AT&T recommends that the Staff consider, at a minimum, revising its recommendation to make it consistent with its recommendations on other issues that require improvement; that is to accept BellSouth's proposal for six months and then require a more stringent interval thereafter.

C-2 Average Arrangement Time

AT&T notes that BellSouth offers shorter intervals for virtual collocation than for physical collocation. For example, in its proposal in Georgia, it offered ordinary physical in 120 calendar days and ordinary virtual in 90 calendar days or 25% less time. Similarly, the Georgia Commission recently ordered[9] ordinary physical in 90 calendar days, and ordinary virtual in 50 calendar days, or 44% less time. AT&T requests that, at a minimum, the Staff revise its recommendation to include a proportionately reduced interval for virtual collocation.

C-3 % of Due Dates Missed

AT&T does not oppose the benchmark of 95% of collocation due dates met.

V. Access to Raw Data and Reporting

Issue 62: Is BellSouth's raw data adequate to validate its reported results?

BellSouth still does not provide raw data for the following CLEC reports:

The six LNP measures
The six billing measures
The trunking measures
The Coordinated Conversions measure

[9] See Interim performance measures established in Docket 8354-U for the third party test.

The Staff's recommendation states that the Commission should allow KPMG to finish its audit of BellSouth's performance measurements, and underlying data, and consider its findings and recommendations. AT&T is not opposed to Staff's recommendation. However, is unclear how an audit will solve the problem of missing raw data files unless KPMG issues an exception, and the Commission requires BellSouth to provide this data. This raw data is currently not provided in Georgia, yet KPMG has issued no exception regarding this deficiency. AT&T therefore requests that the Commission order BellSouth to provide this raw data with which CLECs can validate their reports.

VI. Validation/Audit

Issue 63: Is BellSouth's audit policy as articulated in its SQM appropriate?

AT&T supports the Staff Recommendation and requests that the Staff also recommend that the Commission require BellSouth to modify the audit policy contained in its SQM to comply with the Commission's Order.

VII. Other Issues

Issue 65: What if any access should BellSouth provide to data regarding its performance to its affiliates?

AT&T supports the Staff Recommendation. AT&T recommends that the Commission make clear that this reporting includes all forms service provided to its affiliates to be used in local service, such as interconnection provide to its wireless unit, and the plethora of services provided to its DSL unit, Bellsouth.net.

Issue 66: Should BellSouth's performance to its affiliates become a standard or benchmark where that performance is superior to BellSouth's performance to its retail customers?

AT&T has several concerns with the Staff's recommendations. First, while Staff appeared to agree with AT&T and other CLECs by citing to the FCC rules that an incumbent LEC shall provide a LEC with interconnections " that is at a level of quality that is equal to that which the incumbent LEC provides itself, a subsidiary, an affiliate, or any other party", in its recommendation the Staff only referred to BellSouth CLECs. BellSouth currently provides interconnection to its wireless unit, and DSL services to BellSouth.net. These types of arrangements, as well as any future BellSouth CLEC, must be included to meet the requirements

of the Act and the FCC which used the terminology "itself, a subsidiary, an affiliate, or any other party." As noted in the issue above, the Staff recommendation should make this clear. Second, the Staff's recommendation regarding monitoring and reporting affiliate performance is inconsistent with the FCC's key elements of an enforcement plan:

-a reasonable structure that is designed to detect and sanction poor performance ***when it occurs*** emphasis added)

-a ***self-executing*** mechanism that does not leave the door open unreasonably to litigation and appeal

The Staff's recommendation appears to allow services provided to certain types of affiliates to be excluded from review, and allows poor performance (when compared to the service provided to the affiliate) to remain unsanctioned for months. AT&T requests that the Staff revise its recommendation to correct these deficiencies.

Issue 67: Should BellSouth's Product Interval Guide reflect both mechanized and manual Firm Order Confirmation intervals?

AT&T supports the Staff Recommendation that the Product Interval Guide include the order completion and FOC intervals for mechanized orders.

Issue 68: What structure should be adopted for a remedies plan?

See AT&T's Performance Incentive Plan Version 2.0, August 7, 2000 filing in this docket.

CONCLUSION

For the reasons stated above, in AT&T's Comments filed on March 20, 2000 in this docket, and in the many documents that AT&T placed into the record in this proceeding, AT&T urges the Staff to include in its Recommendation and for the Commission to adopt the positions as set forth by AT&T.

DATED this Seventh day of August, 2000.

By Attorneys:

David L. Guerry
LONG LAW FIRM, L.L.P.
4041 Essen Lane, Suite 500
One United Plaza
Baton Rouge, Louisiana 70809
(225) 922-5110

Gene Coker
AT&T Communications of the South
Central States, Inc.
1200 Peachtree NE
Atlanta, Georgia 30309
(404) 810-8700

Attorneys for AT&T Communications of the
South Central States, Inc.

Certificate

I hereby certify that a copy of the foregoing pleading has been mailed, via U.S. Postage and electronically, to all individuals on the service list of this matter.

Baton Rouge, Louisiana this 7th day of August 2000.

Formatted

David L. Guerry

	Measure	Product Disaggregation	Retail Analog/Benchmark	Other Disaggregation
Provision-ing	Percent Missed Installation Appointments Average Completion Interval % Provisioning Troubles w/I 30 days of Service Order Activity Mean Held Order Interval & Distribution Intervals Average Jeopardy Notice Interval Percentage of Orders Given Jeopardy Notices Average Completion Notice Interval	1. Resale Residence 2. Resale Business 3. Resale Design 4. Resale PBX 5. Resale Centrex 6. Resale ISDN 7. UNE - Analog Loop 8. UNE - Digital Loop < DS1 9. UNE – Digital Loop DS1 10. UNE - Digital Loop > DS1 11. UNE - Switching (ports) 12. UNE - Unbundled Interoffice Transport - Dedicated 13. UNE - Combo (loop + port-analog) 14. UNE-Combo (loop and port-PBX) 15. UNE Combo (loop and port-ISDN) 16. UNE - Combo (other) 17. UNE - xDSL - UL (ADSL, UCL, HDSL) 18. UNE - ISDN - UL 19. UNE - Line Sharing 20. UNE-Sub-loop unbundling 21. Stand-alone LNP 22. Local Interconnection Trunks	1. Parity w retail 2. Parity w retail 3. Parity w retail 4. Parity w retail 5. Parity w retail 6. Parity w retail 7. Retail Res & Bus POTS 8. Retail Digital Loop <DS1 9. Retail DS1 10. Retail >DS1 - 11. Retail POTS 12. Retail DS1 / DS3 – Interoffice 13. Retail POTS 14. Retail PBX 15. Retail ISDN 16. Retail DS1 / DS3 – Interoffice 17. ADSL Provided to Retail 18. Retail ISDN - BRI 19. ADSL Provided to Retail 20. Retail POTS 21. Retail POTS 22. Parity w retail	Dispatch type: None In Out Volume: 1-5 lines 6-14 lines 15+ lines Dispatch type applies to: -Missed Appointment -Order completion Interval -% Provisioning Troubles in 30 days Volume applies to: -Missed Appointment -Order Completion Interval -Held Order

	Measure	Product Disaggregation	Retail Analog/Benchmark	Other Disaggregation
Provision-ing	Average Disconnect Timeliness Interval		95% within 15 minutes of receipt of activate message from NPAC	
Provision-ing	1. % Early Cuts 2. % Troubles within 7 days of a Hot Cut 3. % Late Cuts 4. % of Coordinated Cuts Not Working As Initially Provisioned		1. <5% are provisioned 15 minutes or earlier than due date and time 2. <2% 3. <5% are provisioned outside cutover window 4. <5% are not working when provisioned	
	Measure	Product Disaggregation	Retail Analog/Benchmark	Other Disaggregation

Maintenance & Repair	Missed Repair Appointments Customer Trouble Report Rate Maintenance Average Duration Percent Repeat Troubles w/i 30 days	1. Resale Residence 2. Resale Business 3. Resale Design 4. Resale PBX 5. Resale Centrex 6. Resale ISDN 7. UNE - Analog Loop 8. UNE - Digital Loop < DS1 9. UNE – Digital Loop DS1 10. UNE - Digital Loop >= DS1 11. UNE - Switching (ports) 12. UNE - Unbundled Interoffice Transport - Dedicated 13. UNE - Combo (loop + port-analog) 14. UNE-Combo (loop and port-PBX) 15. UNE Combo (loop and port-ISDN) 16. UNE - Combo (other) 17. UNE - xDSL - UL (ADSL, UCL, HDSL) 18. UNE - ISDN - UL 19. UNE - Line Sharing 20. UNE-Sub-loop unbundling 21. Stand-alone LNP 22. Local Interconnection Trunks	1. Parity w retail 2. Parity w retail 3. Parity w retail 4. Parity w retail 5. Parity w retail 6. Parity w retail 7. Retail Res & Bus POTS 8. Retail Digital Loop <DS1 9. Retail DS1 10. Retail >DS1 11. Retail POTS 12. Retail DS1 / DS3 – Interoffice 13. Retail POTS 14. Retail PBX 15. Retail ISDN 16. Retail DS1 / DS3 - Interoffice 17. ADSL Provided to Retail 18. Retail ISDN - BRI 19. ADSL Provided to Retail 20. Retail POTS 21. Retail POTS 22. Parity w retail	Dispatch Type: --None --In --Out

Statistical Techniques
For The Analysis And Comparison Of
Performance Measurement Data

Submitted to Louisiana Public Service Commission (LPSC)
Docket U-22252 Subdocket C

1. Introduction and Scope

The Louisiana Public Service Commission (LPSC) staff has requested Drs. S. Hinkins, E. Mulrow, and F. Scheuren of Ernst & Young LLP (consultants for BellSouth Telecommunications), and Dr. C. Mallows of AT&T Labs-Research to set out their views on the application of a statistical analysis to performance measurement data. The present report is intended to provide a detailed statistical report on appropriate methodology.

The setting for the analysis is crucial to the interpretation of any statistical significance that might be found. There is no doubt that, to quote the Commission staff, "statistical analysis can help reveal the likelihood that reported differences in an ILECs performance toward its retail customers and CLECs are due to underlying differences in behavior rather than random chance" (Staff Final Recommendation, LPSC Docket No. U-22252 - Subdocket C, dated August 12, 1998, pages 15 - 16).

To frame our presentation the next paragraph from the LPSC Docket U-22252 is quoted in its entirety.

> **"Statistical tests are effective in identifying those measurements where differences in performance exist. The tests themselves cannot identify the cause of the apparent differences. The differences may be due to a variety of reasons, including: 1) when the ILEC and CLEC processes being measured are actually different and should not be expected to produce the same result, 2) when the ILEC is employing discriminatory practices, or 3) when assumptions necessary for the statistical test to be valid are not being met."** (Ibid., page 16)

Apparent statistically significant differences in BellSouth and CLEC performance can arise when

- the ILEC and CLEC processes being measured are actually different and should not be expected to produce the same result
- the ILEC is employing discriminatory practices, or
- assumptions necessary for the statistical test to be valid are not being met.

To meet the Louisiana Commission's purpose, we will recommend techniques that are robust in the presence of possible assumption failure, carefully examine BellSouth Telecommunications (BST) and CLEC performance so "like" is compared only to "like,"

and are still able, in a highly efficient manner, to detect differences. Upon investigation any differences detected might lead to concerns about possible discriminatory practices.

The LPSC staff also states "that a uniform methodology which identifies those items which need to be measured, how they are to be measured, and how the results are to be reported is also desirable and would be beneficial to all parties" (Ibid., page 16). We agree with this goal as well, stipulating only that the use of a single method may not be desirable while a single methodology (or a set of methods) could be.

The statistical process for testing if CLEC and ILEC customers are being treated equally involves more than just a mathematical formula. Three key elements need to be considered before an appropriate decision process can be developed. These are

- the type of data,
- the type of comparison, and
- the type of performance measure.

When examining the various combinations of these elements, we find that there is a set of testing principles that can be applied uniformly. However, the statistical formulae that need to be used change as the situation changes.

To be responsive to the Commission, we have divided our discussion into four sections and five appendices. The contents of each of these are briefly mentioned below -- first for the main report and then for the extensive supporting appendix materials.

For the main report, this section (Section I) introduces our work and sets out the required scope. The next two sections (Sections II and III) discuss the type of comparisons that need to be identified, and the appropriate testing principles. The final section (Section IV) provides an overview of appropriate testing methodologies, based on what we have learned from our examination of BellSouth's performance measure data in Louisiana.

The five appendices provide technical details on the statistical calculations involved in the Truncated Z statistic (Appendix A), the implementation of the methodology for the trunk blocking performance measure (Appendix B), the calculations involved in computing the balancing critical value of a test (Appendix C), ways to present the results using detailed statistical displays so that results can be audited (Appendix D), and the technical details involved in data trimming (Appendix E).

2. Data Considerations, Comparisons, and Measurement Types

This section makes general distinctions which apply to the performance measures. These distinctions will be important in the determination of appropriate methodologies.

Data Set Types. The type of statistical methodology used depends on the form of the data available. In general, there are two ways to classify the data used for performance measure comparisons. These are:

- transaction level data, and
- aggregated summaries.

Records in a transaction level data set represent a single transaction, e.g. an individual customer order, or the record of a specific trouble reported by a customer. This type of data set allows for deep like-to-like comparisons, and may also allow one to identify the root cause of a problem. A testing methodology needs to be carefully chosen so that it incorporates the comparison levels and does not cover up problem areas.

Records in an aggregated summary data set are typically summaries of related transactions. For example, the total number of blocked calls in a trunk group during the noon hour of a day is a summary statistic. This type of data set may not contain as much information as a transaction level data set, and it therefore needs to be treated differently. While a general methodology may be determined for a transaction level data set, it may not be possible to do so for aggregated summaries. Testing methodology needs to be developed on a case-by-case basis.

Comparison Types. An ILEC's performance in providing services to CLEC customers is tested in one of two ways:

- by comparing CLEC performance to ILEC performance when a retail analog exists, or
- by comparing CLEC performance to a benchmark.

The testing methodologies for these two situations will have similarities, but there are differences that need to be understood.

Table 1 categorizes those performance measures that E&Y has examined by data type and comparison type. The table shows that five performance measures with retail analogs have transaction level data, while three others with retail analogs only have summary level data. No performance measures using benchmarks have been studied.

Table 1. Classification of Performance Measures by Data and Comparison Type (only measures previously examined by E&Y are included)

Level of Data	Comparison Type	
	Retail Analog	**Benchmark**
Transaction Level	Order Completion Interval Maintenance Average Duration % Missed Installations % Missed Repair Trouble Report Rate	No Measures Examined
Summary Level	Billing Timeliness OSS Response Interval Trunk Blocking	No Measures Examined

Measurement Types. The performance measures that will undergo testing are of three types: means, proportions (an average of a measure that takes on only the values of 0 or 1), and rates.

While all three have similar characteristics, proportions and rates are derived from count data while a mean is an average of interval measurements. Table 2 classifies the performance measures by the type of measurement.

Table 2: Classification of Performance Measures by Measurement Type

Mean	Proportion	Rate
Order Completion Interval Maint. Ave. Duration OSS Response Interval	Percent Missed Installations Percent Missed Repairs Billing Timeliness Trunk Blocking	Trouble Report Rate

3. Testing Principles

This section describes five general principles which the final methodology should satisfy:

1. *When possible, data should be compared at appropriate levels, e.g. wire center, time of month, dispatched, residential, new orders.*
2. *Each performance measure of interest should be summarized by one overall test statistic giving the decision maker a rule that determines whether a statistically significant difference exists.*
3. *The decision system must be developed so that it does not require intermediate manual intervention.*
4. *The testing methodology should balance Type I and Type II Error probabilities.*
5. *Trimming of extreme observations from BellSouth and CLEC distributions is needed in order to ensure that a fair comparison is made between performance measures.*

Like-to-Like Comparisons. *When possible, data should be compared at appropriate levels, e.g. wire center, time of month, dispatched, residential, new orders.*

In particular, to meet this goal the testing process should:

- Identify variables that may affect the performance measure.
- Record important confounding covariates.
- Adjust for the observed covariates in order to remove potential biases and to make the CLEC and the ILEC units as comparable as possible.

It is a well know principle that comparisons should be made on equal footing: apples-to-apples, oranges-to-oranges. Statistical techniques that are addressed in most text books usually assume that this is the case beforehand. Some higher level books address the issue of "designed experiments" and discuss appropriate ways to structure the data collection method so that the text books' formulae can be used in analyzing the data.

Performance measure testing does not involve data from a designed experiment. Rather, the data is obtained from an observational study. That being the case, one must impose a structure on the data after it is gathered in order to assure that fair comparisons are being made. For example, it is important to disaggregate the data to a fine level so that appropriate like-to-like comparisons of CLEC and ILEC data can be made. Any statistical methodology that ignores important confounding variables can produce biased results.

Aggregate Level Test Statistic. *Each performance measure of interest should be summarized by one overall test statistic giving the decision maker a rule that determines whether a statistically significant difference exists.*

To achieve this goal, the aggregate test statistic should have the following properties:

- The method should provide a single overall index, on a standard scale.

- If entries in comparison cells are exactly proportional over a covariate, the aggregated index should be very nearly the same as if comparisons on the covariate had not been done.
- The contribution of each comparison cell should depend on the number of observations in the cell.
- Cancellation between comparison cells should be limited, i.e., positive outcomes should not be allowed to cancel negative ones.
- The index should be a continuous function of the observations.

Since the data are being disaggregated to a very deep level, thousands of like-to-like comparison cells are created. It would be an extremely laborious task for a decision maker to sort through individual test results for each cell and determine if discrimination exists. An aggregate summary statistic is needed in order to make an overall judgment.

The aggregate level statistic should be insensitive to small changes in cells values, and its value should not be affected if some of the disaggregation for like-to-like cells is truly unnecessary. Furthermore, individual cell results should be weighted so that those cells with more transactions have larger effects on the overall result.

Production Mode Process. *The decision system must be developed so that it does not require intermediate manual intervention.*

Two statistical paradigms are possible for examining performance measure data. In the exploratory paradigm, data are examined and methodology is developed that is consistent with what is found. In a production paradigm a methodology is decided upon before data exploration. For the production paradigm to succeed

- Calculations should be well defined for possible eventualities.
- The decision process should be based on an algorithm that needs no manual intervention.
- Results should be arrived at in a timely manner.
- The system must recognize that resources are needed for other performance measure-related processes that also must be run in a timely manner.
- The system should be both auditable and adjustable over time.

While the exploratory paradigm provides protection against using erroneous data, it requires a great deal of lead time and is unsuitable for timely monthly performance measure testing. A production paradigm will not only promptly produce overall test results but will also provide documentation that can be used to explore the data after the test results are released.

Error Probability Balancing. *The testing methodology should balance Type I and Type II Error probabilities.*

Specifically, what is required to achieve this goal is

- The probability of a Type I error should equal the probability of a Type II error for well-defined null and alternative hypotheses.
- The formula for a test's balancing critical value should be simple enough to calculate using standard mathematical functions, i.e. one should avoid methods that require computationally intensive techniques.
- Little to no information beyond the null hypothesis, the alternative hypothesis, and the number of observations should be required for calculating the balancing critical value.

The objective of a statistical test is to test a hypothesis concerning the values of one or more population parameters. Usually an inquiry into whether or not there is evidence to support a hypothesis, called the *alternative hypothesis*, is conducted by seeking statistical evidence that the converse of the alternative, the *null hypothesis*, is most likely false. If there is not sufficient evidence to reject the null hypothesis, then a case for accepting the alternative has not been made.

Two types of errors are possible in any decision-making process. These have been summarized in Table 3.

Table 3: Statistical Testing Errors

Decision Error	General Description	In terms of Performance Measure Testing
Type I	Rejecting the null hypothesis (accepting the alternative) when the null is true.	Deciding that BST favors its own customers when it does not.
Type II	Accepting the null hypothesis when the alternative is true.	Deciding that BST does not favor its own customers when it does.

In a controlled experimental study where the sample sizes are relatively small, it is generally desirable to control the Type I error closely to avoid making a conclusion that there is a difference when, in fact, there is none. The probability of a Type II error is not directly controlled but is determined by the sample size and the distance between the null and the alternative hypotheses. Thus, there is some kind of balance between Type I and Type II errors with Type I error usually controlled more closely.

If a standard of materiality is set by stating a specific alternative for the test, and the distribution of the test statistic under both the null and alternative hypotheses is understood, then a critical value can be determined so that the two error probabilities are equal.

Trimming. *Trimming of extreme observations from BellSouth and CLEC distributions is needed in order to ensure that a fair comparison is made between performance measures.*

Three conditions are needed to accomplish this goal. These are:

- Trimming should be based on a general rule that can be used in a production setting.
- Trimmed observations should not simply be discarded; they need to be examined and possibly used in the final decision making process.
- Trimming should only be used on performance measures that are sensitive to "outliers."

For the purpose of performance measure testing, trimming refers to removing transactions that significantly distort the performance measure statistic for the set of transactions under consideration. For example, the arithmetic average (or mean) is extremely sensitive to "outliers" since a single large value can significantly distort the average.

The term "outliers" refers to:

1) extreme data values that may be valid, but since they are rare measurements, they may be considered to be statistically unique; or
2) large values that should not be in the analysis data set because of errors in the measurement or in selecting the data.

Trimming is beneficial since it puts both ILEC and CLEC transactions on equal footing with respect to the largest value in each set. Note, though, that it is only needed for performance measures that are distorted by outliers. Of the three types of measures defined in Section 2, only mean (average) measures require trimming. Appendix E sets forth a trimming plan for mean performance measures.

4. Testing Methodology

This section details the testing methodology that is most appropriate for the various types of performance measures. First, transaction level testing will be discussed when there is a retail analog. Next, transaction level testing against a benchmark. Then, testing when only aggregated summaries are available.

Transaction Level - Retail Analog: The Truncated Z Statistic. When a retail analog is available CLEC performance can be directly compared with ILEC performance. Over the last year, for transaction level data, many test statistics have been examined. We now believe that the "Truncated Z" test statistic provides the best compromise with respect to possessing the desired qualities outlined in Section 3, above.

The Truncated Z is fully described in Appendix A, and formulae for calculation of a balancing critical value are found in Appendix C. The main features of this statistic are:

- A basic test statistic is calculated within each comparison cell.
- The value of a cell's result is left "as is" if the result suggests that "favoritism" may be taking place. Otherwise, the result is set to zero. This is called the truncation step.
- Weights that depend on the volume of both ILEC and CLEC transactions within the cell are determined, and a weighted sum of the "truncated" cell results is calculated.
- The weighted sum is theoretically corrected to account for the truncation, and a final overall statistic is determined.
- This overall test value is compared to a balancing critical value to determine if favoritism is likely.

The test statistic itself is based on like-to-like comparisons, and it possesses all five of the properties of an aggregate test statistic (Section 3). While the test requires a large amount of calculations, our studies of the process on some of BellSouth's performance measure data indicate that the calculations can be completed in a reasonable amount of time. Therefore, the process can be put into production mode. Finally, since a balancing critical value can be calculated, it is possible to balance the error probabilities.

Transaction Level - Benchmark. When a benchmark is used, CLEC performance is not compared with ILEC performance. Like-to-like comparison cells are not needed, thus greatly simplifying the testing process. Statistical testing can be done using a probability model, or non-statistical testing can be done using a deterministic model. No data for this data/comparison class has been studied at this point in time.

If one wants a method that is independent of the number of transactions, then statistical methods should be used to determine if observed performance below the benchmark is statistically significant. Once again, we want a procedure that adheres to the principles outlined in Section 3.

Aggregated Summary - Retail Analog or Benchmark. We cannot provide any one single set of rules for the analysis of data in this class. Data that is an aggregated summary of transactions may or may not present problems. For example, BellSouth's trunk blocking data is saved as summaries by hour of the day. Collectively, the summaries provide sufficient information to proceed with the Truncated Z methodology.

On the other hand, our examination of the data for the OSS response interval revealed that information necessary for computing a Truncated Z was not available. In this case, however, we were able to construct a satisfactory time series method to analyze the measure.

Each measure falling into this class needs to be handled on a case-by-case basis. If sufficient information is available to use the Truncated Z method, then we feel it should

be used. When the Truncated Z cannot be used, a testing methodology that adheres closely to the principles outlined in Section 3 should be determined and followed.

Appendix A. The Truncated Z Statistic

The Truncated Z test statistic was developed by Dr. Mallows in order to have an aggregate level test when transaction level data are available that

- provides a single overall index on a standard scale;
- is robust with respect to unnecessary disaggregation,
- incorporates the number of observations in a cell into the determination of the weight for the contribution of each comparison cell,
- limits the amount of "neutralization" between comparison cells, and
- is a continuous function of the observations.

The Ernst & Young statistical team and Dr. Mallows have studied the implementation of the statistic using some of BellSouth's performance measure data. This has resulted in an overall process for comparing CLEC an ILEC performance such that the following principles hold:

1) Like-to-Like Comparisons are made. (See Appendix B for an example based on the trunk blocking measure.)
2) Error probabilities are balanced. (See Appendix C)
3) Extreme values are trimmed from the data sets when they significantly distort the performance measure statistic. (See Appendix E)
4) The testing process is an automated production system. (Discussed here. See Appendix D for reporting guidelines.)
5) The determination of ILEC favoritism is based on a single aggregate level test statistic. (Discussed here.)

This appendix provides the details behind computing the Truncated Z test statistic so that principles 4 and 5 hold. We start by assuming that any necessary trimming of the data is complete, and that the data are disaggregated so that comparisons are made within appropriate classes or adjustment cells that define "like" observations.

Notation and Exact Testing Distributions

Below, we have detailed the basic notation for the construction of the truncated z statistic. In what follows the word "cell" should be taken to mean a like-to-like comparison cell that has both one (or more) ILEC observation and one (or more) CLEC observation.

L = the total number of occupied cells

j = 1,...,L; an index for the cells

n_{1j} = the number of ILEC transactions in cell j

n_{2j} = the number of CLEC transactions in cell j

n_j = the total number transactions in cell j; $n_{1j}+ n_{2j}$

X_{1jk} = individual ILEC transactions in cell j; k = 1,..., n_{1j}

X_{2jk} = individual CLEC transactions in cell j; k = 1,..., n_{2j}

Y_{jk} = individual transaction (both ILEC and CLEC) in cell j

$$= \begin{cases} X_{1jk} & k = 1,\ldots,n_{1j} \\ X_{2jk} & k = n_{1j}+1,\ldots,n_j \end{cases}$$

$\Phi^{-1}(\cdot)$ = the inverse of the cumulative standard normal distribution function

For Mean Performance Measures the following additional notation is needed.

$\overline{X}_{1j}$ = the ILEC sample mean of cell j

$\overline{X}_{2j}$ = the CLEC sample mean of cell j

s^2_{1j} = the ILEC sample variance in cell j

s^2_{2j} = the CLEC sample variance in cell j

y_{jk} = a random sample of size n_{2j} from the set of $Y_{j1},\ldots,Y_{jn_j}$; k = 1,...,n_{2j}

M_j = the total number of distinct pairs of samples of size n_{1j} and n_{2j};

$$= \binom{n_j}{n_{1j}}$$

The exact parity test is the permutation test based on the "modified Z" statistic. For large samples, we can avoid permutation calculations since this statistic will be normal (or Student's t) to a good approximation. For small samples, where we cannot avoid permutation calculations, we have found that the difference between "modified Z" and the textbook "pooled Z" is negligible. We therefore propose to use the permutation test based on pooled Z for small samples. This decision speeds up the permutation computations considerably, because for each permutation we need only compute the sum of the CLEC sample values, and not the pooled statistic itself.

A permutation probability mass function distribution for cell j, based on the "pooled Z" can be written as

$$PM(t) = P(\sum_k y_{jk} = t) = \frac{\textit{the number of samples that sum to } t}{M_j},$$

and the corresponding cumulative permutation distribution is

$$CPM(t) = P(\sum_{k} y_{jk} \le t) = \frac{\textit{the number of samples with sum } \le t}{M_j}.$$

For Proportion Performance Measures the following notation is defined

a_{1j} = the number of ILEC cases possessing an attribute of interest in cell j

a_{2j} = the number of CLEC cases possessing an attribute of interest in cell j

a_j = the number of cases possessing an attribute of interest in cell j; $a_{1j} + a_{2j}$

The exact distribution for a parity test is the hypergeometric distribution. The hypergeometric probability mass function distribution for cell j is

$$HG(h) = P(H = h) = \begin{cases} \dfrac{\binom{n_{1j}}{h}\binom{n_{2j}}{a_j - h}}{\binom{n_j}{a_j}}, & \max(0, a_j - n_{2j}) \le h \le \min(a_j, n_{1j}) \\ 0 & \text{otherwise} \end{cases},$$

and the cumulative hypergeometric distribution is

$$CHG(x) = P(H \le x) = \begin{cases} 0 & x < \max(0, a_j - n_{1j}) \\ \sum_{h=\max(0, a_j - n_{1j})}^{x} HG(h), & \max(0, a_j - n_{1j}) \le x \le \min(a_j, n_{2j}) \\ 1 & x > \min(a_j, n_{2j}) \end{cases}.$$

For Rate Measures, the notation needed is defined as

b_{1j} = the number of ILEC base elements in cell j

b_{2j} = the number of CLEC base elements in cell j

b_j = the total number of base elements in cell j; $b_{1j} + b_{2j}$

$\hat{r}_{1j}$ = the ILEC sample rate of cell j; n_{1j}/b_{1j}

$\hat{r}_{2j}$ = the CLEC sample rate of cell j; n_{2j}/b_{2j}

q_j = the relative proportion of CLEC elements for cell j; b_{2j}/b_j

The exact distribution for a parity test is the binomial distribution. The binomial probability mass function distribution for cell j is

$$BN(k) = P(B = k) = \begin{cases} \binom{n_j}{k} q_j^k (1-q_j)^{n_j-k}, & 0 \le k \le n_j \\ 0 & \text{otherwise} \end{cases},$$

and the cumulative binomial distribution is

$$CBN(x) = P(B \le x) = \begin{cases} 0 & x < 0 \\ \sum_{k=0}^{x} BN(k), & 0 \le x \le n_j \\ 1 & x > n_j \end{cases}.$$

Calculating the Truncated Z

The general methodology for calculating an aggregate level test statistic is outlined below.

1. **Calculate cell weights**, W_j. A weight based on the number of transactions is used so that a cell which has a larger number of transactions has a larger weight. The actual weight formulae will depend on the type of measure.

   Mean Measure

   $$W_j = \sqrt{\frac{n_{1j} n_{2j}}{n_j}}$$

   Proportion Measure

   $$W_j = \sqrt{\frac{n_{2j} n_{1j}}{n_j} \cdot \frac{a_j}{n_j} \cdot \left(1 - \frac{a_j}{n_j}\right)}$$

   Rate Measure

   $$W_j = \sqrt{\frac{b_{1j} b_{2j}}{b_1} \cdot \frac{n_j}{b_j}}$$

2. **In each cell, calculate a Z value**, Z_j. A standard normal Z statistic is needed for each cell.

   - If $W_j = 0$, set $Z_j = 0$.
   - When the cell sample sizes are sufficiently large, formulae based on a normal approximation can be used.
   - If cell sample sizes are not large enough for a normal approximation to hold, then exact testing methods must be employed. When this occurs, the results of the test statistic are converted into an equivalent value from the standard normal distribution.

   The actual Z statistic calculation depends on the type of performance measure.

   Mean Measure

$$Z_j = \Phi^{-1}(\alpha)$$

   where α is determine by the following algorithm.

   If $\min(n_{1j}, n_{2j}) > 6$, then determine α as

$$\alpha = P(t_{n_{1j}-1} \le T_j),$$

   that is, α is the probability that a t random variable,

$$t_j = \frac{\bar{X}_{1j} - \bar{X}_{2j}}{s_{1j}\sqrt{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}}$$

   with n_{1j} - 1 degrees of freedom, is less than T_j.

$$T_j = t_j + \frac{g}{6}\left(\frac{n_{1j} + 2n_{2j}}{\sqrt{n_{1j}\, n_{2j}(n_{1j} + n_{2j})}}\right)\left(t^2 + \frac{n_{2j} - n_{1j}}{2n_{1j} + n_{2j}}\right).$$

   Here the coefficient g is an estimate of the skewness of the parent population, which we assume is the same in all cells. It can be estimated from the ILEC values in the largest cells. This needs to be done only once for each measure. We have found that attempting to estimate this skewness parameter for each cell separately leads to excessive variability in the "adjusted" t. We therefore use a single compromise value in all cells.

   Note, that t_j is the "modified Z" statistic. The statistic T_j is a "modified Z" corrected for the skewness of the ILEC data.

If $\min(n_{1j}, n_{2j}) \leq 6$, and

a) $M_j \leq 1{,}000$ (the total number of distinct pairs of samples of size n_{1j} and n_{2j} is 1,000 or less).

- Calculate the sample sum for all possible samples of size n_{2j}.
- Rank the sample sums from smallest to largest. Ties are dealt by using average ranks.
- Let R_0 be the rank of the observed sample sum with respect all the sample sums.

$$\alpha = 1 - \frac{R_0 - 0.5}{M_j}$$

b) $M_j > 1{,}000$

- Draw a random sample of 1,000 sample sums from the permutation distribution.
- Add the observed sample sum to the list. There is a total of 1001 sample sums. Rank the sample sums from smallest to largest. Ties are dealt by using average ranks.
- Let R_0 be the rank of the observed sample sum with respect all the sample sums.

$$\alpha = 1 - \frac{R_0 - 0.5}{1001}.$$

Proportion Measure

$$Z_j = \frac{n_j a_{1j} - n_{1j} a_j}{\sqrt{\frac{n_{1j} n_{2j} a_j (n_j - a_j)}{n_j - 1}}}.$$

Rate Measure

If $n_{1j}>15$, $n_{2j}>15$, and $n_j q_j (1-q_j) > 9$ then

$$Z_j = \frac{\hat{r}_{1j} - \hat{r}_{2j}}{\sqrt{\hat{r}_{1j}\left(\frac{1}{b_{1j}} + \frac{1}{b_{2j}}\right)}}.$$

Otherwise,

$$Z_j = \Phi^{-1}(\alpha)$$

where

$$\alpha = 1 - CBN(n_{2j}-1).$$

3. **Obtain a truncated Z value for each cell,** Z_j^*. To limit the amount of cancellation that takes place between cell results during aggregation, cells whose results suggest possible favoritism are left alone. Otherwise the cell statistic is set to zero. This means that positive equivalent Z values are set to 0, and negative values are left alone. Mathematically, this is written as

$$Z_j^* = \min(0, Z_j).$$

4. **Calculate the theoretical mean and variance of the truncated statistic under the null hypothesis of parity,** $E(Z_j^* | H_0)$ and $Var(Z_j^* | H_0)$. In order to compensate for the truncation in step 3, an aggregated, weighted sum of the Z_j^* will need to be centered and scaled properly so that the final aggregate statistic follows a standard normal distribution.

   - If $W_j = 0$, then no evidence of favoritism is contained in the cell. The formulae for calculating $E(Z_j^* | H_0)$ and $Var(Z_j^* | H_0)$ cannot be used. Set both equal to 0.

   - If the equivalent Z value of a mean or rate measure was calculated using a normal approximation, or $\min(n_{1j}, n_{2j}) > 30$ for a proportion measure then

$$E(Z_j^* | H_0) = -\frac{1}{\sqrt{2\pi}}, \text{ and}$$

$$Var(Z_j^* | H_0) = \frac{1}{2} - \frac{1}{2\pi}.$$

   - Otherwise, determine the total number of values for Z_j^*, denoted by N_j. Let z_{ji} and θ_{ji}, $i = 1,\ldots,N_j$, denote the values of Z_j^* and the probabilities of observing each value, respectively.

$$E(Z_j^* | H_0) = \sum_{i=1}^{N_j} \theta_{ji} z_{ji}, \text{and}$$

$$\mathrm{Var}(Z_j^* \mid H_0) = \sum_{i=1}^{N_j} \theta_{ji} z_{ji}^2 - \left[E(Z_j^* \mid H_0) \right]^2 .$$

The actual values of the z's and θ's depends on the type of measure.

Mean Measure

$$N_j = \min(M_j, 1{,}000)$$

$$z_{ji} = \min\left\{0, 1 - \Phi^{-1}\left(\frac{R_i - 0.5}{N_j}\right)\right\} \quad \text{where } R_i \text{ is the rank of sample sum i}$$

$$\theta_j = \frac{1}{N_j}$$

Proportion Measure

$$N_j = \min(a_j, n_{2j}) - \max(0, a_j - n_{1j}) + 1$$

$$z_{ji} = \min\left\{0, \frac{n_j\, i - n_{1j}\, a_j}{\sqrt{\frac{n_{1j}\, n_{2j}\, a_j\, (n_j - a_j)}{n_j - 1}}}\right\}, \quad i = 0, \ldots, a_j$$

$$\theta_{ji} = \mathrm{HG}(i)$$

Rate Measure

$$N_j = n_j$$

$$z_{ji} = \min\{0, \Phi^{-1}(1 - \mathrm{CBN}(i-1))\}$$

$$\theta_{ji} = \mathrm{BN}(i)$$

5. **Calculate the aggregate test statistic,** Z^T.

$$Z^T = \frac{\sum_j W_j Z_j^* - \sum_j W_j E(Z_j^* \mid H_0)}{\sqrt{\sum_j W_j^2 \mathrm{Var}(Z_j^* \mid H_0)}}$$

Decision Process

Once Z^T has been calculated, it is compared to a critical value to determine if the ILEC is favoring its own customers over a CLEC's customers. The derivation of the critical value is found in Appendix C.

This critical value changes as the ILEC and CLEC transaction volume change. One way to make this transparent to the decision maker, is to report the difference between the test statistic and the critical value, $diff = Z^T - c_B$. If favoritism is concluded when $Z^T < c_B$, then the $diff < 0$ indicates favoritism.

This make it very easy to determine favoritism: a positive *diff* suggests no favoritism, and a negative *diff* suggests favoritism. Appendix D provides an example of how this information can be reported for each month.

Appendix B. Trunk Blocking

This Appendix describes how the trunk blocking data can be processed to apply the Truncated Z Statistic. Trunk blocking is defined as the proportion of blocked calls a trunk group experiences in a time interval. It is a ratio of two numbers—blocked and attempted calls, both of which can vary over time and across trunk groups. Since the measure is a proportion where the numerator is a subset of the denominator, the truncated Z statistic, modified for proportions, can be applied here (see Appendix A).

As with other performance measures, data are first assigned to like-to-like cells, and the Z statistic is then computed within each cell. For trunk blocking, cells are defined by three variables: hour, day, and trunk group size or capacity. The next sections will describe the data and the data processing steps in greater detail.

Data Sources

Two data files are processed for the trunk blocking measure. One is the Trunk Group Data File that contains the Trunk Group Serial Number (TGSN), Common Language Location Identifier (CLLI) , and other characteristics needed to categorize trunk groups and to identify them as BellSouth or CLEC.

The other file is the Blocking Data File (BDF), which contains the actual 24 hour blocking ratios for each weekday. There are 4 or 5 weeks in a monthly report cycle. The current system, however, allows the storage of daily blocking data by hour for a week only. Therefore, the data elements necessary to compute the Truncated Z must be extracted each week.

Two important data fields of interest on the Blocking Data File are the Blocking Ratio and Offered Load. The basic definition of Blocking Ratio is the proportion of all attempted calls that were blocked. For the simplest case of one way trunk groups, this is computed by dividing the number of blocked calls by the total call attempts, given that the data are valid. If they are not valid (e.g., actual usage exceeds capacity), blocking is estimated via the Neal Wilkinson algorithm.

Although the raw data--blocked calls (overflow) and peg counts (total call attempts)--are available, the calculation of the Blocking Ratio may be complicated for two-way trunk groups and trunk groups with invalid data. For this reason, we use the blocking ratios from the BDF instead of computing the ratios from the raw data. In order to reflect different call volumes processed through each trunk group, however, the blocking ratios need to be either weighted by call volume or converted to blocked and attempted calls before they are aggregated.

The measure of call traffic volume recommended for weighting is Offered Load. Offered Load is different from call counts in that it incorporates call duration as well. Since it is

not just the number of calls but the total usage—number of calls multiplied by average call duration--that determines the occurrence of any blocking, this pseudo measure, Offered Load, appears to be the best indicator of call volume.

Cells or comparison classes are determined by three factors—hour, day, and trunk group capacity (number of trunks in service). The first two factors represent natural classes because trunk blocking changes over time. The third factor is based on our finding that high blocking tends to occur in small trunk groups. A pattern was found not only in the magnitude of blocking but also in its variability. Both the magnitude and variability of blocking decrease as trunk group capacity increases. Additional work is needed to establish the appropriate number of capacity levels and the proper location of boundaries.

Data Processing

The data are processed using the five steps below:

1. Merge the two files by TGSN and select only trunk groups listed in both files.
2. Reset the blocking of all high use trunk groups to zero[1].
3. Assign trunk group categories to CLEC and BellSouth: Categories 1, 3, 4, 5, 10, and 16 for CLEC and 9 for BellSouth[2]. The categories used here for comparison are:

Category	Administrator	Point A	Point B
1	BellSouth	BellSouth End Office	BellSouth Access Tandem
3	BellSouth	BellSouth End Office	CLEC Switch
4	BellSouth	BellSouth Local Tandem	CLEC Switch
5	BellSouth	BellSouth Access Tandem	CLEC Switch
9	BellSouth	BellSouth End Office	BellSouth End Office
10	BellSouth	BellSouth End Office	BellSouth Local Tandem
16	BellSouth	BellSouth Tandem	BellSouth Tandem

4. Recode the missing data. The Blocking Data File assigns all missing data (no valid measurement data) zero blocking. To differentiate true zero blocking from zeroes due to missing data, invalid records were identified and the ratios reset to missing. The blocking value was invalid if both the number of Loaded Days and the Offered Load were 0 for a given hourly period.
5. Form comparison classes based either on the data (i.e., quartiles) or on a predetermined set of values.

[1] The high use trunk groups cannot have any blocking. These are set up such that all overflow calls are automatically routed to other trunk groups instead of being physically blocked.
[2] More detailed information on all categories is described in a report 'Trunk Performance Report Generation' by Ernst & Young (March 1999).

Calculation of the Proportion of Blocked Calls

Each cell is determined by day of the month, hour of the day, and trunk group capacity. To use the Truncated Z method, we generate summary information, to include the total number of blocked calls and the total number of attempted calls, for each cell.

For the details of each calculation step, the following notation is used. For a given hour of a day, let $\overline{X}_{1ij}$ be the proportion of BellSouth blocked calls for trunk group i in cell j and $\overline{X}_{2ij}$ be the corresponding proportion for CLEC. Then $\overline{X}_{1ij} = X_{1ij} / n_{1ij}$ where X_{1ij} denotes the number of BellSouth blocked calls and n_{1ij} denotes the number of BellSouth total call attempts (indicated by Offered Load) for trunk group i in cell j. Likewise, $\overline{X}_{2ij} = X_{2ij} / n_{2ij}$. For the steps outlined below, only the CLEC notation is provided.

1. Compute the number of blocked calls for trunk group i: $X_{2ij} = \overline{X}_{2ij} * n_{2ij}$
2. Compute total call attempts for all trunk groups in the cell: $n_{2j} = \sum_i n_{2ij}$
3. Compute mean blocking proportion for cell j: $\overline{X}_{2j} = \sum_i X_{2ij} \Big/ \sum_i n_{2ij}$
4. Compute the total number of BellSouth and CLEC blocked calls in cell j: $t_j = \sum_i X_{1ij} + \sum_i X_{2ij}$
5. Apply the Truncated Z Statistic for Proportion measures presented in Appendix A.

Appendix C
Balancing the Type I and Type II Error Probabilities of the Truncated Z Test Statistic

This appendix describes a the methodology for balancing the error probabilities when the Truncated Z statistic, described in Appendix A, 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 Truncated Z test statistic, Z^T, and
4. a critical value, c

The decision rule[1] is

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

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

Type I Error: Deciding favoritism exists when there is, in fact, no favoritism.
Type II Error: Deciding parity exists when there is, in fact, favoritism.

The probabilities of each type of each are:

Type I Error: $\alpha = P(Z^T < c \mid H_0)$.
Type II Error: $\beta = P(Z^T \geq c \mid H_a)$.

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

[1] This decision rule assumes that the smaller a performance measure is, the better the service. If the opposite is true, then reverse the decision rule.

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

where

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

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

$SE(\hat{T}|H_0)$ 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 a standard normal distribution. Here

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

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

Notice that

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

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

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

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

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

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

$$c_B = \frac{E(\hat{T}|H_a) - E(\hat{T}|H_0)}{SE(\hat{T}|H_a) + SE(\hat{T}|H_0)} \qquad (C.4)$$

The Balancing Critical Value of the Truncated Z

In Appendix A, the Truncated Z statistic is defined as

$$Z^T = \frac{\sum_j W_j Z_j^* - \sum_j W_j E(Z_j^*|H_0)}{\sqrt{\sum_j W_j^2 Var(Z_j^*|H_0)}}$$

In terms of equation (C.1) we have

$$\hat{T} = \sum_j W_j Z_j^*$$

$$E(\hat{T}|H_0) = \sum_j W_j E(Z_j^*|H_0)$$

$$SE(\hat{T}|H_0) = \sqrt{\sum_j W_j^2 Var(Z_j^*|H_0)}$$

To compute the balancing critical value (C.4), we also need $E(\hat{T}|H_a)$ and $SE(\hat{T}|H_a)$. These values are determined by

$$E(\hat{T}|H_a) = \sum_j W_j E(Z_j^*|H_a), \text{ and}$$

$$SE(\hat{T}|H_a) = \sqrt{\sum_j W_j^2 \, var(Z_j^*|H_a)}.$$

In which case equation (C.4) gives

$$c_B = \frac{\sum_j W_j E(Z_j^*|H_a) - \sum_j W_j E(Z_j^*|H_0)}{\sqrt{\sum_j W_j^2 \, var(Z_j^*|H_a)} + \sqrt{\sum_j W_j^2 \, var(Z_j^*|H_0)}}. \qquad (C.5)$$

Thus, we need to determine how to calculate $E(Z_j^*|H_0)$, $Var(Z_j^*|H_0)$, $E(Z_j^*|H_a)$, and $Var(Z_j^*|H_a)$. These values depend on the distribution of Z_j (see Appendix A) under the null and alternative hypotheses.

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

H_0: $\mu_{1j} = \mu_{2j}$, $\sigma_{1j}^2 = \sigma_{2j}^2$

H_a: $\mu_{2j} = \mu_{1j} + \delta_j \cdot \sigma_{1j}$, $\sigma_{2j}^2 = \lambda_j \cdot \sigma_{1j}^2$ $\quad \delta_j > 0$, $\lambda_j \geq 1$ and $j = 1,\ldots,L$.

Under this null hypothesis, Z_j has a standard normal distribution within each cell j. In which case,

$$E(Z_j^* | H_0) = -\frac{1}{\sqrt{2\pi}}, \text{ and}$$
$$\text{var}(Z_j^* | H_0) = \frac{1}{2} - \frac{1}{2\pi}.$$

Under the alternative hypothesis, Z_j has a normal distribution with

$$E(Z_j | H_a) = m_j = \frac{-\delta_j}{\sqrt{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}}, \text{ and}$$

$$SE(Z_j | H_a) = se_j = \sqrt{\frac{\lambda_j n_{1j} + n_{2j}}{n_{1j} + n_{2j}}}$$

In general, the mean of a normal distribution truncated at 0 is

$$M(\mu,\sigma) = \int_{-\infty}^{0} \frac{x}{\sqrt{2\pi\sigma}} \exp(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2) dx,$$

and the variance is

$$V(\mu,\sigma) = \int_{-\infty}^{0} \frac{x^2}{\sqrt{2\pi\sigma}} \exp(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2) dx - M(\mu,\sigma)^2$$

It can be shown that

$$M(\mu,\sigma) = \mu\,\Phi(\tfrac{-\mu}{\sigma}) - \sigma\,\phi(\tfrac{-\mu}{\sigma})$$

and

$$V(\mu,\sigma) = (\mu^2 + \sigma^2)\Phi(\tfrac{-\mu}{\sigma}) - \mu\,\sigma\,\phi(\tfrac{-\mu}{\sigma}) - M(\mu,\sigma)^2$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function, and $\phi(\cdot)$ is the standard normal density function.

Using the above notation, and equation (C.5), we get the formula for the balancing critical of Z^T for the alternative hypothesis defined above.

$$c_B = \frac{\sum_j W_j M(m_j, se_j) - \sum_j W_j \frac{-1}{\sqrt{2\pi}}}{\sqrt{\sum_j W_j^2 V(m_j, se_j)} + \sqrt{\sum_j W_j^2 \left(\frac{1}{2} - \frac{1}{2\pi}\right)}}. \qquad \text{(C.6)}$$

This formula assumes that Z_j, is approximately normally distributed within cell j. When the cell sample sizes, n_{1j} and n_{2j}, are small this may not be true. It is possible to determine the cell mean and variance under the null hypothesis when the cell sample sizes are small. It is much more difficult to determine these values under the alternative hypothesis. Since the cell weight, W_j will also be small (see Appendix A) for a cell with small volume, the cell mean and variance will not contribute much to the weighted sum. Therefore, formula (C.6) should provide a reasonable approximation to the balancing critical value.

Determining the Parameters of the Alternative Hypothesis

In this appendix we have indexed the alternative hypothesis by two sets of parameters, λ_j and δ_j. 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 Choices for λ_j. The set of parameters λ_j index alternatives to the null hypothesis that arise 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. While concerns about differences in the variability of service are important, it turns out that the truncated Z testing which is being recommended here is relatively insensitive to all but very large values of the λ_j. Put another way, reasonable differences in the values chosen here could make very little difference in the balancing points chosen.

- Parameter Choices for δ_j. The set of parameters δ_j are much more important in the choice of the balancing point than was true for the λ_j. The reason for this is that they directly index differences in average service. The truncated Z test is very sensitive to any such differences; hence, even small disagreements among experts in the choice of the δ_j could be very important. Sample size matters here too. For example, setting all the δ_j to a single value — $\delta_j = \delta$ — might be fine for tests across individual CLECs where currently in Louisiana the CLEC customer bases are not too different. Using the same

value of δ for the overall state testing does not seem sensible, however, since the state sample would be so much larger.

The bottom line here is that beyond a few general considerations, like those given above, a principled approach to the choice of the alternative hypotheses to guard against must come from elsewhere.

Appendix D: Reporting Guidelines

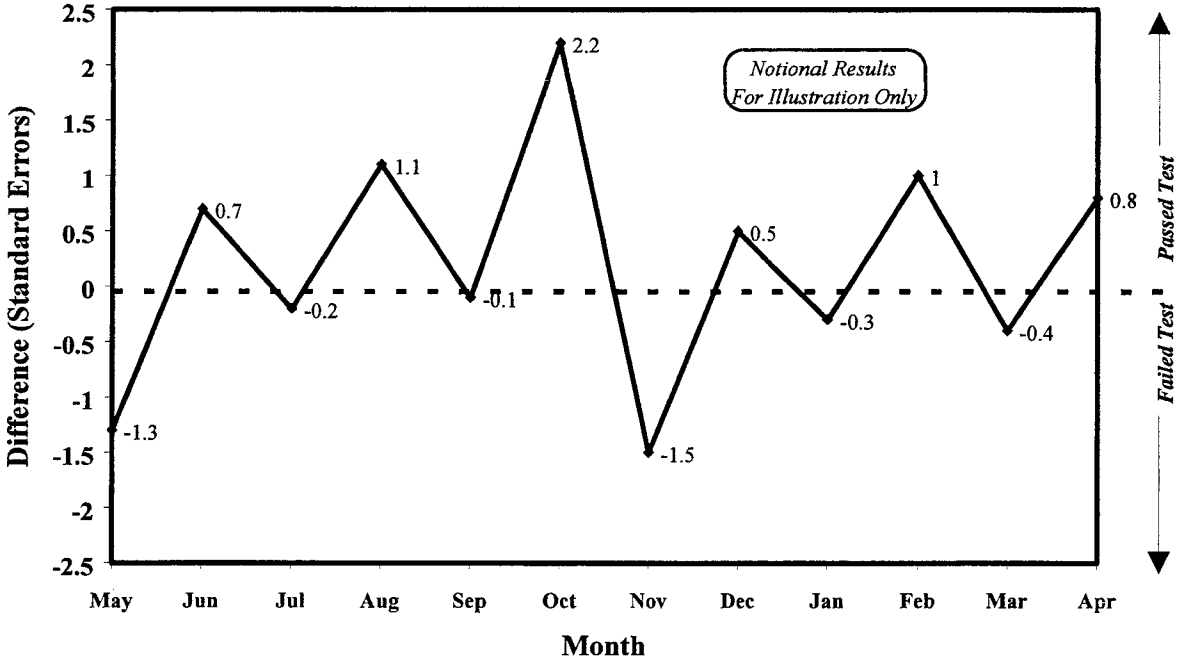
The general structure for reporting statistical results in a production environment will be the same for the different measures and will consist of three components. For each measure, we will first present the monthly test statistics over a period of time. Following this will be the results for the current month, with summary statistics, test statistics, and descriptive graphs. Finally, the third component of the reporting structure is a summary of any adjustments to the data made in the process of running the tests, mainly including a description of how many records were excluded from analysis and the reason for the exclusion (i.e., excluded due to business rules, or due to statistical/methodological rules pertaining to the measure). This component is important to assure that the reported results can be audited.

Selected components of the reporting structure are illustrated in the samples that follow. An outline of the report is shown below. Monthly results will be presented for each level of aggregation required.

I. Test Statistics Over Time
II. Monthly Results
    A. Summary Statistics
    B. Test Statistics
    C. Descriptive Graphs (Frequency Distributions, etc.)
III. Adjustments to Data
    A. Records Excluded Due to Business Rules
    B. Records Excluded Due to Statistical Rules

Test Statistic Over Time. The first component of the reporting structure is an illustration of the trend of the particular performance measure over time together with a tabular summary of results for the current month. We will show at a glance whether the tests consistently return non-statistically significant results; consistently indicate disparity (be that in favor of BellSouth or in favor of the CLECs); or vary month by month in their results. An example of this component follows.

Notional Performance Measure
Through April XXXX
Differences Between Test Statistic and Balancing Critical Value

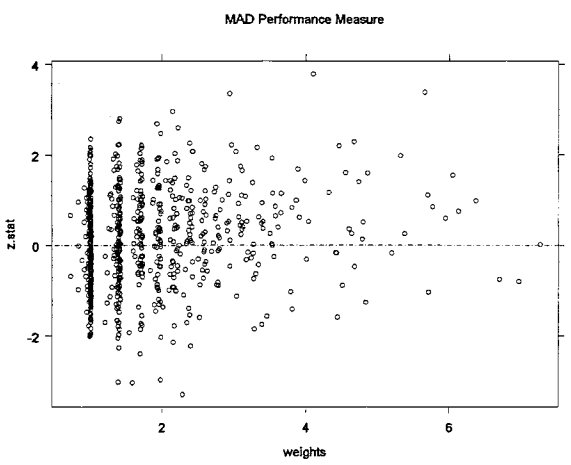


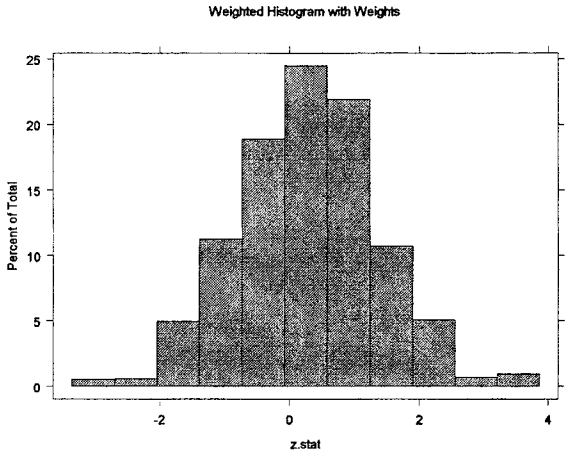
Result for Current Month	
Test Statistic	0.410
Balancing Critical Value	-1.210
Difference	0.800

Monthly Results. The most important component of the reporting structure is the part which presents results of the monthly statistical tests on the given performance measure. The essential aspects included in this component are the summary statistics; the test statistics and results; and descriptive graphs of the results.

It is important to present basic summary statistics to complete the comparison between BellSouth and the CLECs. At a minimum, these statistics will include the means, standard deviations, and population sizes. In addition to basic descriptive statistics, we also present the test statistic results. Examples of ways we have presented these statistics in the past can be found in BellSouth's February 25, 1999 filing before the Louisiana Public Service Commission.

Finally, the results will be presented in graphical format. Below is an example of how to graphically present the data behind the Truncated Z statistic. One graph shows a plot of cell Z score versus cell weights. The other is a histogram of the weighted cell Z scores.





Adjustments to Data. The third important component of the reporting structure is information on any adjustments performed on the data. This information is essential in order that the results may be verified and audited. The most prevalent examples of such modifications would be removal of observations and weighting of the data.

Records can be removed from analysis for both business reasons (these will likely be taken into account in the PMAP system) and for statistical reasons. All of the performance measures exclude certain records based on business rules underlying each measure's particular definitions and methodologies. The number of records excluded for each rule will be summarized. In addition, some of the measures will have observations excluded for statistical reasons, particularly in the case of "mean measures" (OCI and MAD); these exclusions will be summarized as well. The tables below show examples of the current method for summarizing this information:

April XXXX
Perormance Measure Filtering Information

This table displays information about the size of the database files and the cases that were removed from the analysis.

CLEC		1999	BST	
Unfiltered Total	**28,691**		**Unfiltered Total**	**453,107**
Records Removed for Business Reasons *(e.g. not N, T, C, or P orders, not resale and not UNE)*	**7,242**		**Records Removed for Business Reasons** *(e.g. not N, T, C, or P orders, not retail)*	**78,613**
Total Reported on Web Report	**21,449**		**Total Reported on Web Report**	**374,494**
Additional Records Removed for Business Reasons	**876**		**Additional Records Removed for Business Reasons**	**7,429**
Missing Appointment code is 'S'	844		Missing Appointment code is 'S'	7,172
General Class Service = 'O'	0		General Class Service = 'O'	279
UNE Cases	102			
Records Removed for Statistical Reasons Extreme Values Removed	**9**		**Records Removed for Statistical Reasons Extreme Values Removed**	**652**
No Matching Classification Removals	**47**		**No Matching Classification Removals**	**21,974**
FILTERED TOTAL	**20,517**		**FILTERED TOTAL**	**344,439**

Appendix E. Trimming Outliers for Mean Measures

The arithmetic average is extremely sensitive to outliers; a single large value, possibly an erroneous value, can significantly distort the mean value. And by inflating the error variance, this also affects conclusions in the test of hypotheses. Extreme data values may be correct, but since they are rare measurements, they may be considered to be statistical outliers. Or they may be values that should not be in the analysis data set because of errors in the measurement or in selecting the data.

At this time, only two mean measures have been analyzed: Order Completion Interval and Maintenance Average Duration. Maintenance Average Duration data are truncated at 240 hours and therefore this measure was not trimmed further. For Order Completion Interval, the underlying distribution of the observations is clearly not normal, but rather skewed with a very long upper-tail.

A useful technique, coming from the field of robust statistical analysis, is to trim a very small proportion from the tails of the distribution before calculating the means. The resulting mean is referred to as a trimmed mean. Trimming is beneficial in that it speeds the convergence of the distribution of the means to a normal distribution. Only extreme values are trimmed, and in many cases the data being trimmed are, in fact, data that might not be used in the analysis on other grounds.

In the first analysis of the verified Order Completion Interval-Provisioning measure, after removing data that were clearly in error or were not applicable, we looked at the cases that represented the largest 0.01% of the BST distribution. In the August data, this corresponded to orders with completion intervals greater than 99 days. All of these were BellSouth orders. In examining the largest 11 individual examples that would be removed from analysis, we found that only 1 of the 11 cases was a valid case where the completion interval was unusually large. The other 10 cases were examples of cases that should not have been included in the analysis. This indicates that at least in preliminary analysis, it is both beneficial to examine the extreme outliers and reasonable to remove them.

A very slight trimming is needed in order to put the central limit theorem argument on firm ground. But finding a robust rule that can be used in a production setting is difficult. Also, any trimming rule should be fully explained and any observations that are trimmed from the data must be fully documented.

When it is determined that a measure should be trimmed, a trimming rule that is easy to implement in a production setting is:

> **Trim the ILEC observations to the largest CLEC value from all CLEC observations in the month under consideration.**

That is, no CLEC values are removed; all ILEC observations greater than the largest CLEC observation are trimmed.

While this method is simple, it does allow for extreme CLEC observations to be part of the analysis. For instance, suppose that the amount of time to complete an order was less than 40 days for all CLEC orders except one. Let's say that this extreme order took 100 days to complete. The trimming rule says that all ILEC orders above 100 days should be trimmed, but a closer look at the data might suggest trimming at 40 days instead.

Since we are operating in a production mode system, it is not possible to explore the data before the trimming takes place. Other automatic trimming rules present other problems, so our solution is to use the simple trimming rule above, and have the system automatically produce a trimming report that can be examined at a later point in time.

The trimming report should include:

- The value of the trim point.
- Summary statistics and graphics of the ILEC observations that were trimmed.
- A listing of the trimmed ILEC transaction for a random sample of 10 trimmed transactions. This listing should not disclose sensitive information.
- A listing of the 10 most extreme CLEC transactions. This listing should not disclose sensitive information.
- The number of ILEC and CLEC observations above some fixed point, so that changes in the upper tail can be better tracked over time.

The trimming report should be part of the overall report discussed in Appendix D. Examples of tables contained within the trimming report are shown below.

April XXXX
Performance Measure Extreme Values

	CLEC		BST
Cutoff	26	Cutoff	26
# of Records	20,573	# of Records	367,065
10 Largest		Extreme Values	652
Minimum	19	Minimum	27
Median	23	Median	32
Maximum	26	Maximum	283
Subtotal	**20,573**	**Subtotal**	**366,413**

April XXXX
Performance Measure Weighting Report

	CLEC		BST
# of Records	20,573	# of Records	366,413
No Matching BST		No Matching CLEC	
Classification (1)	47	Classification (2)	21,974
Subtotal	**20,526**	**Subtotal**	**344,439**

April XXXX
Perormance Measure Filtering Information

This table displays information about the size of the database files and the cases that were removed from the analysis.

CLEC		1999	RST	
Unfiltered Total	**28,691**		**Unfiltered Total**	**453,107**
Records Removed for Business Reasons *(e.g. not N, T, C, or P orders, not resale and not UNE)*	**7,242**		**Records Removed for Business Reasons** *(e.g. not N, T, C, or P orders, not retail)*	**78,613**
Total Reported on Web Report	**21,449**		**Total Reported on Web Report**	**374,494**
Additional Records Removed for Business Reasons	**876**		**Additional Records Removed for Business Reasons**	**7,429**
Missing Appointment code is 'S'	844		Missing Appointment code is 'S'	7,172
General Class Service = 'O'	0		General Class Service = 'O'	279
UNE Cases	102			
Records Removed for Statistical Reasons			**Records Removed for Statistical Reasons**	
Extreme Values Removed	**0**		**Extreme Values Removed**	**652**
No Matching Classification Removals	**47**		**No Matching Classification Removals**	**21,974**
FILTERED TOTAL	**20,526**		**FILTERED TOTAL**	**344,439**

CLEC Extreme Values

Wire Center	Time	Dispatch	Residence	Circuits	Order Type	Order Interval
NWORLAMA	1	1	3	1	N	61
OPLSLATL	1	2	1	1	C	53
NWORLAMA	2	1	3	1	N	44
NWORLAMA	1	1	3	1	N	39
BTRGLAWN	1	1	2	1	C	38
LKCHLADT	1	1	1	1	T	37
NWORLAMA	1	1	3	1	N	32
NWORLAMA	2	1	3	1	N	32
SHPTLACL	1	1	2	1	N	28

Frequency of Extreme Values Removed from BST file (Top 10)

Wire Center	Time	Dispatch	Residence	Circuits	Order Type	Frequency
NWORLAMA	1	1	3	1	N	55
NWORLAMA	2	1	3	1	N	25
BTRGLASB	2	1	3	1	C	23
NWORLAMC	2	1	3	1	C	23
NWORLAMC	1	1	3	1	C	22
NWORLAMA	2	1	3	1	C	18
NWORLAMA	1	1	3	1	C	17
BTRGLASB	1	1	3	1	C	16
LFYTLAMA	1	1	3	1	C	15
NWORLAMA	2	2	3	1	C	14

Attachment 3

Revised language for Section 4.4.7.2 of SEEM plan

Attachment 3

Response will be provided as soon as possible.

Attachment 4

Additional language for Section 4.6.1 of SEEM plan

4. Please provide a sentence which could be added to Change of Law section 4.6.1 to achieve the intent of your alternative language.

If a change of law occurs which may affect BellSouth's obligations to CLECs, CLECs may Petition the Commission within 30 days to seek changes to the SQM and SEEM plans in accordance with such change of law.

Attachment 5

Comments regarding the feasibility of using parity test results at the CLEC aggregate level, etc.

5. Please provide comments regarding the feasibility of using parity test results at the CLEC aggregate level, by sub-measure and disaggregating or allocating the aggregate affected volume to individual CLECs in place of proposed language for 4.3.1.2 regarding small sample sizes.

The CLECs believe that aggregating Bell South wholesale performance for the CLEC by sub-measure and then disaggregating or allocating the aggregate affected volume to individual CLECs in place of Bell South proposed language for section 4.3.1.2 regarding small sample size, presents a poor choice for the CLECs. Furthermore, the SEEM, as written, provides a statistical methodology for analyzing parity standard metrics with small sample sizes. This methodology was originally agreed upon by both Bell South and the CLECs, was ordered by the Florida Commission (as well as other state Commissions), and represents a scientifically valid basis for detecting discrimination, when it occurs, for parity comparisons based on small sample sizes.

As indicated in the CLEC initial response to the Bell South proposed change for section 4.3.1.2, the CLECs disagree with the Bell South proposal that the minimum number of transactions should be increased to 30 per CLEC per measure to generate a SEEM calculation. The CLECs also disagree that cells with less than 5 transactions should be excluded from the calculations. The current SEEM structure consisting of the BCV methodology coupled with the truncated Z-test was originally formulated and agreed upon by Bell South and CLEC statisticians for use in Louisiana and much of the region.[1] Furthermore, with the addition of procedures for exact non-parametric analysis (i.e., permutation testing or its equivalent) small sample size comparisons may be accurately made.[2] [3] For example, Triola states:

> "...But given another experiment with a relatively small amount of data drawn from some mysterious population (that is, one for which we didn't know the distribution), we would probably fare better with a nonparametric test. Sometimes we don't really have a choice."[4]

---

[1] Statistical Techniques For The Analysis And Comparison Of Performance Measurement Data Submitted to Louisiana Public Service Commission (LPSC) Docket U-22252 Subdocket C, 1999.

[2] *Ibid*, Appendix A, The Truncated Z Statistic, page A-2.

[3] Florida Seem Administrative Plan, Version 4.01, p27-29 (See technical attachment to this action item.).

[4] Mario F. Triola, **Elementary Statistics 6e**, Addison-Wesley Publishing Company, Reading, Massachusetts, 1995, pp 654.

The approporiate method for accommodating possible small sample sizes that may arise from CLEC-level disaggregation is the method first suggested by the statistical studies of the experts in Louisiana and subsequently ordered by the Florida commission:

> "ORDERED that the Truncated Z statistic shall be used to evaluate compliance for enforcement measures with retail analogs. For small samples (30 or less), a permutation test shall be used to calculate Z-scores for mean measures. ... For small samples, the hypergeometric test, also known as Fisher's Exact Test, shall be used for proportion and rate measures."[5]

Also, note the Commission's comments before ordering the above methodology:

> "For small samples (30 or less), BellSouth witness Mulrow, ALEC Coalition witness Bell, and [Z]-Tel witness Ford agree that a permutation test should be used to calculate Zscores for mean measures."[6]

> "We find that there will be little practical difference in the range of values for these two types of measures, in the context of a performance assessment plan. For small samples, all witnesses who offered an opinion stated that the hypergeometric test, also known as Fisher's Exact Test, is appropriate for proportion and rate measures."[7] [8]

Thus, the current SEEM methodology is based upon extensive theoretical and empirical study and consequently designed to accommodate the full range of sample sizes and cell occupancies. Furthermore, the SEEM methodology, formulated with current sample size minimums, has operated for many years and under many performance conditions and localities, including large and small transaction counts, CLECs, states, and under temporal (seasonal) variation. Moreover, Bell South has indicated that benchmark standard should take account, but not ignore, small "sample size" results by using a special table when CLEC transaction counts are between five and thirty. Permutation analysis is

---

[5] ORDER NO. PSC-01-1819-FOF-TP, DOCKET NO. 000121-TP, ISSUED: September 10, 2001, PAGE 201 – 202.

[6] ORDER NO. PSC-01-1819-FOF-TP, DOCKET NO. 000121-TP, ISSUED: September 10, 2001, PAGE 151.

[7] ORDER NO. PSC-01-1819-FOF-TP, DOCKET NO. 000121-TP, ISSUED: September 10, 2001, page 152.

[8] Actually the theoretically correct distribution to be used for rate measures is the Binomial., which the Florida SEEM structure requires (See technical attachment to this action item.). However, as long as the rates being measured are small compared to 100%, the difference between the hypergeometric and binomial distributions is negligible, as noted by the Commission order.

the analogous current procedure in SEEM for analyzing parity standard measures with small sample size. Bell South has neither demonstrated how their proposed modified sample size minima come about, nor how accuracy of the SEEM plan is increased with their proposed modifications. In fact, detection of discrimination, where it exists, will necessarily be reduced by the Bell South proposals of removing potentially discriminatory performance from SEEM consideration. The CLECs consider these proposed modifications as a vehicle solely for Bell South to reduce its liability under the SEEM plan.

The suggested modification to aggregate CLEC data, by sub-measure, each month, in order to increase testing sample size assumes that the current statistical methodology in the SEEM is somehow inappropriate or inaccurate, however, as we have seen, it is viable and reasonable. Furthermore, the CLECs are very concerned that the aggregation process would degrade implementation/operation information by effectively averaging wholesale performance over all the individual CLECs; thereby potentially masking discrimination. For example, suppose within a wire center CLEC A operates with a relatively high wholesale mean time-to-repair but with a small number of troubles. CLEC A's performance will be un- or less-noticeable if aggregated with data from CLEC B, which has a larger number of troubles, but a smaller mean time-to-repair. If larger CLEC B has the benefit of a more efficient account team than the smaller CLEC A, then CLEC B may have parity performance (or better) with Bell South while CLEC A might not. Two troublesome results can easily ensue. First, CLEC B data may completely dominate that of CLEC A. In this case a statistical test will not find significant difference between Bell South and the aggregated CLEC (A + B) data; therefore no remedies will ensue to CLEC A. Second, if CLEC B also has sufficiently poor wholesale performance that the statistical test finds a significant difference between the CLEC aggregate and Bell South, the affected volume per CLEC may not be accurately determined because parity differences between Bell South and CLEC A or CLEC B may not be proportional to their respective sample sizes.[9] Therefore, if CLEC A's poor performance is due to discrimination by Bell South's wholesale operations, not only is there a strong possibility that the discrimination would go unnoticed, but the SEEM remedies may not accurately reflect what should be paid and to whom.

[9] We also note that aggregation can cause Bell South to pay remedies to CLEC A, with which it may be in parity, if CLEC B has less than parity performance. Such eventualities decrease plan accuracy

Technical Attachment for CLEC Action Item 5

C.2.1 Mean Measures

For mean measures, an adjusted, asymmetric t statistic is calculated for each like-to-like cell that has at least seven BST and seven CLEC transactions. A permutation test is used when one or both of the BST and CLEC sample sizes is less than seven. The adjusted, asymmetric t statistic and the permutation calculation are described in Appendix D, Statistical Formulas and Technical Description.

C.2.2 Proportion Measures

For performance measures that are calculated as a proportion, in each adjustment cell, the cell Z and the moments for the truncated cell Z can be calculated in a direct manner. In adjustment cells where proportions are not close to zero or one, and where the sample sizes are reasonably large ($n_{ij}p_{ij}(1-p_{ij}) > 9$), a normal approximation can be used. In this case, the moments for the truncated Z come directly from properties of the standard normal distribution. If the normal approximation is not appropriate, then the Z statistic is calculated from the hypergeometric distribution. In this case, the moments of the truncated Z are calculated exactly using the hypergeometric probabilities.

C.2.3 Rate Measures

The truncated Z methodology for rate measures has the same general structure for calculating the Z in each cell as proportion measures. For the rate measure customer trouble report rate there are a fixed number of access lines in service for the CLEC, b_{2j}, and a fixed number for BST, b_{1j}. The modeling assumption is that the occurrence of a trouble is independent between access lines, and the number of troubles in b access lines follows a Poisson distribution with mean λ_b where λ is the probability of a trouble per 1 access line and b (= $b_{1j} + b_{2j}$) is the total number of access lines in service. The exact permutation distribution for this situation is the binomial distribution (the limit for the hypergeometric distribution) that is based on the total number of BST and CLEC troubles, n, and the proportion of BST access lines in service, $q_j = b_{1j}/b$.

In an adjustment cell, if the number of CLEC troubles is greater than 15 and the number of BST troubles is greater than 15, and $n_{ij}q_{ij}(1-q_{ij}) > 9$, then a normal approximation can be used. In this case, the moments of the truncated Z come directly from properties of the standard normal

distribution. Otherwise, if there are very few troubles, the number of CLEC troubles can be modeled using a binomial distribution with n equal to the total number of troubles (CLEC plus BST troubles.) In this case, the moments for the truncated Z are calculated explicitly using the binomial distribution.