AFFIDAVIT

DISTRICT OF COLUMBIA

BEFORE ME, the undersigned authority, duly commissioned and qualified in and for the District aforesaid, personally came and appeared Edward J. Mulrow, BellSouth Telecommunications, Inc., being by me first duly sworn deposed and said that:

He is appearing as a witness before the Kentucky Public Service Commission in "Investigation Concerning the Propriety of InterLATA Services by BellSouth Telecommunications, Inc. Pursuant to the Telecommunications Act of 1996," KY PSC Case No. 2001-105, and if present before the Commission and duly sworn, his testimony would be set forth in the annexed transcript consisting of 27 pages and 0 exhibit(s).

Edward J. Mulrow

SWORN TO AND SUBSCRIBED BEFORE ME this
23rd day of July, 2001.

Laura L. Montgomery
NOTARY PUBLIC

LAURA L. MONTGOMERY
Notary Public, District of Columbia
My Commission Expires Nov. 30, 2005

BELLSOUTH TELECOMMUNICATIONS, INC.

REBUTTAL TESTIMONY OF EDWARD J. MULROW, PH.D.

BEFORE THE KENTUCKY PUBLIC SERVICE COMMISSION
DOCKET NO. 2001-105

JULY 30, 2001

Q. PLEASE STATE YOUR NAME, AND BUSINESS NAME AND ADDRESS.

A. My name is Edward J. Mulrow. I am employed by Ernst & Young LLP as a Senior Manager in the Quantitative Economics and Statistics Group. I have been retained by BellSouth as a statistical advisor. My business address is 1225 Connecticut Ave., NW, Washington, DC 20036.

Q. ARE YOU THE SAME EDWARD J. MULROW THAT FILED DIRECT TESTIMONY IN THIS DOCKET?

A. Yes. I filed direct testimony in this docket on May 18, 2001.

Q. WHAT IS THE PURPOSE OF YOUR REBUTTAL TESTIMONY?

A. The purpose of my rebuttal testimony is to respond to portions of the rebuttal testimony of Dr. Robert M. Bell representing AT&T Communications of the Southern States. In responding to the rebuttal testimony of this witness, I address the following issues:

- Conclusions regarding KCI's third-party test statistical procedures.
- The appropriate statistical methodology for making performance measure

parity comparisons in an enforcement plan such as BellSouth's SEEM penalty plan.

- Dr. Bell's analysis of the impact of "delta."
- The appropriateness of BellSouth's SEEM remedy calculation.

Q. WOULD YOU PLEASE GIVE A GENERAL SUMMARY OF DR. BELL'S ASSESSMENT OF THE STATISTICAL METHODOLOGY THAT KCI USED IN THE GEORGIA OSS THIRD PARTY TEST?

A. Dr. Bell's main argument appears to be that KPMG Consulting (KCI) should have used the same principles in statistically testing performance data as were used to develop the statistical methods used in penalty plans such as BellSouth's SEEM plan.

Q. DO ALL STATISTICAL COMPARISONS OF PERFORMANCE DATA NEED TO BE BASED ON THE SAME PRINCIPLES?

A. No. When developing a statistical methodology for any problem, one needs to consider the purpose of the testing that will be done, and the way in which data will be collected. OSS testing is designed for different purposes than enforcement testing such as the statistical methods used in SEEM. Therefore one would not necessarily use the same principles and methods in the testing. In their Motion for Leave to the Louisiana Public Service Commission, KCI

states that their intention was to use statistics as information in their overall evaluation of BellSouth's OSS; that is statistics were not the sole determining factor as they are in enforcement plans.

Q. DO YOU BELIEVE THAT THE STATISTICAL APPROACH KCI USED IN THE GEORGIA THIRD PARTY OSS TESTING WAS APPROPRIATE?

A. I have not been involved in OSS third party testing, so I do not have first hand knowledge to judge if KCI's approach is appropriate. BellSouth witness Mr. Pate testifies that KCI used a similar approach in the New York OSS third party test, and that this approach is consistent with that of Telcordia in the Texas OSS third party test. The methods used in these OSS tests were found to be appropriate by the FCC, so I conclude that KCI had good reason to use the statistical methodology it did, even though it is based on different principles than the SEEM plan.

Q. DR. ROBERT BELL PROPOSES THAT THE KENTUCKY COMMISION ORDER THE MODIFIED Z AS A COMPONENT OF THE STATISTICAL METHODOLOGY. DO YOU AGREE?

A. Yes, Dr. Bell and I agree on this point. The Truncated Z statistic, which BellSouth incorporates into its SEEM plan, is an aggregation of modified Z statistics. And based in his testimony in this docket (page 31, lines 17 – 20), I

conclude that Dr. Bell agrees that if a penalty plans allows for a reasonable and appropriate level of aggregation, the use of the Truncated Z, as BellSouth has proposed, is appropriate.

Q. IF THERE IS NO DISAGREEMENT ON THE TEST STATISTIC THAT SHOULD BE USED IN AN ENFORCEMENT PLAN, THEN WHAT, IF ANY, DIFFERENCES EXIST IN THE STATISTICAL PARTS OF THE PROPOSED PLANS?

A. The differences that exist in the statistical parts of the plans are not due to competing statistical methodologies. The differences lie in the important decisions that need to be made in order to carry out the statistical tests. In my direct testimony, I mention that the Truncated Z methodology needs to have performance data broken into like-to-like categories, which we refer to as cells. The AT&T performance plan (given in AT&T witness Ms. Bursh's testimony) also calls for a disaggregation into like-to-like categories, but AT&T advocates a different disaggregation than BellSouth.

Also, AT&T does not want to re-aggregate the results of the statistical comparisons. Instead, they want decisions regarding parity to be made for each individual test statistic that is calculated. BellSouth on the other hand has proposed groupings for cell results to be re-aggregated using the Truncated Z statistic in order to arrive at a smaller set of global decisions on parity of service.

Q. DO THE JUDGMENTS REGARDING DISAGGREGATION AND STATISTICAL RE-AGGREGATION REST IN THE HANDS OF THE STATISTICIANS?

A. No, decisions regarding the appropriate disaggregation of transactions, and the reasonable levels of re-aggregation should primarily be based on business judgment. This does not mean that statisticians have no role in the process. The impact of the many choices that can be made need to be understood by the decision makers. Statisticians play an important role in describing this impact, but in the end, the decisions are best left in the hands of telecommunications experts and regulators.

Q. DO YOU HAVE ANY INDICATION AS TO WHETHER OR NOT DR. BELL AGREES WITH YOU ON THIS ISSUE?

A. Yes. Once again if we review Dr. Bell's testimony in recent Florida and North Carolina hearings on performance measure issues, we see that he takes a similar position. (For example, see "Investigation Into The Establishment Of Operations Support Systems Permanent Performance Measures For Incumbent Local Exchange Telecommunications Companies." Florida Public Service Commission, Docket No. 000121-TP, volume 6, Cross examination of Dr. Robert Michael Bell, page 1097, lines 2 – 23.)

Q. IS BELLSOUTH THE ONLY COMPANY SUGGESTING THAT SOME FORM OF STATISTICAL AGGREGATION BE DONE?

A. No. The five states where the FCC has granted an RBOC the right to market long distance services have performance comparison plans that aggregate the results of many comparisons into an overall result that determines parity/disparity.

In New York and Massachusetts, Verizon uses a weighted average of performance scores to make parity judgments. In Texas, Oklahoma, and Kansas, Southwestern Bell uses the "K-value" method. This "K-value" methodology is described by AT&T's Dr. Mallows in the "Affidavit of Dr. Colin L. Mallows before the Federal Communications Commission" (sworn May 29, 1998). Thus, both of the methods of aggregation that AT&T's expert has suggested have been adopted by former Bell Companies for use in their performance plans. AT&T however, appears reluctant to accept either of these methodologies.

Q. ARE THERE OTHER DIFFERENCES BETWEEN WHAT DR. BELL HAS PROPOSED AND THE BELLSOUTH METHODOLOGY?

A. Yes. While we both agree that a "modified" z statistic should be used for like-to-like level comparisons, the formulae that Dr. Bell proposes for testing proportion and rate measures are slightly different than those given in Louisiana Statistician's Report (attachment EJM-1 of my direct testimony). In order to explain the root of the problem, however, I need to tell you something more about statistics. The basic modified Z statistic is a comparison of

"means," that is, we take the average of the BellSouth transactions in the "cell" and compare that "average" or "mean" to the comparable "mean" of the CLEC transactions. Not all observations lend themselves to the calculation of "means," however. For instance, consider "missed appointments." With "missed appointments" you are looking at the percentage of the total number of scheduled appointments that were missed. As a result, you end up with a proportion, such as a tenth of a percent or 5 percent or whatever figure is appropriate. You do not have a mean per se. Another example is what we call a "rate" such as the "customer trouble report rate", where you are looking at the number of troubles BellSouth or the CLEC has per the number of available lines. Unlike the "proportional" measures described above, which would always have to be less than 1, the measurement of a "rate" could exceed 100 percent. For instance, if you had ten access lines and 12 reported troubles (that is some lines have more than a single trouble during the reporting period) you can get more than a figure of 100 percent. Again, these two special categories are simply different measures than the "means" calculation that we have been talking about.

The root of the problem is that Dr. Bell uses the modified Z concept irrespective of whether the measure is one based on "means," "proportions," or "rates." The difficulty from a statistical perspective is that the concept the modified Z statistic is based on should not be applied across the board to all measure types. Specifically, the basis for the modified Z statistic is that you

take the difference between the two "means" in the particular "cell" or sub-measure, and divide the result by the standard deviation of BellSouth's mean. This is done to make the test sensitive to changes in the CLEC standard deviation (compared to the BellSouth standard deviation) that would be harmful to the CLEC. In other words, BellSouth could try to give the same average service to CLEC customers as to it own customers, but do so in a way that some CLEC customers receive longer completion times. For example, suppose that BellSouth always services its own customers in 2 days. BellSouth could service one-third of the CLEC customers in 1 day, one-third in two days, and the remaining third in 3 days. On the average, the CLEC service times are the same as BellSouth's, but one-third of the CLEC customers received service that was "below" average. Dividing the difference between the means by only BellSouth's standard deviation avoids masking this problem.

The same situations cannot occur for "proportion" or "rate" measures. In the case of a proportion, such as "missed appointments" that is stated as a percentage of total appointments scheduled, you only have one parameter to consider, the proportionality. As a result, BellSouth cannot separately control the proportion value and the variability about that value.

Q. ARE THERE DIFFERENT FORMULAS THAT DR. BELL COULD HAVE USED TO ADDRESS THESE ISSUES?

A. Yes, the formulae for proportion and rate measures that are given in the Louisiana Statistician's Report are appropriate. Dr. Bell still relies on the outdated LCUG version 1.0 document (attachment RMB-1 of Dr. Bell's direct testimony) for the definition of the test statistics. The methodology described in LCUG version 1.0 was developed in a vacuum devoid of real performance measurement data. All those involved in the analysis of BellSouth performance measure for the Louisiana report, including Dr. Mallows, another AT&T statistician, carefully considered the appropriate formulae to use for proportion and rate measures, and concluded that the LCUG version 1.0 formulae were not appropriate.

Q. THERE IS ANOTHER IMPORTANT INPUT PARAMETER THAT IS NEEDED FOR THE BALANCING METHODOLOGY THAT BOTH BELLSOUTH AND AT&T AGREE TO USE, NAMELY "DELTA." IS THE CHOICE OF "DELTA" ALSO BASED ON BUSINESS JUDGMENT?

A. Yes. As I stated in my direct testimony, 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 should be made based on economic/business judgment.

Q. IN ORDER TO SUPPORT HIS CHOICE OF A "DELTA" VALUE OF 0.25, DR. BELL PROVIDES A TABLE SHOWING THE PERCENTAGE OF CLEC CUSTOMERS RECEIVING BAD SERVICE, BY BELLSOUTH PERCENT AND DELTA. CAN YOU COMMENT ON THIS TABLE?

A. The table you are referring to is Table 2 on page 36 of Dr. Bell's rebuttal testimony. This table is based on a proportion measure, and BellSouth does not use "delta" to define the alternative hypothesis for proportion measures.

Q. DOES DR. BELL UNDERSTAND THAT BELLSOUTH DOES NOT USE "DELTA" FOR PROPORTION MEASURES?

A. Yes, I believe he does. At a Florida hearing, in response to a question of whether or not a table very similar to Table 2 represents what BellSouth is proposing for proportion measures, Dr. Bell said, "It does not represent what they are proposing for proportion measures." (See "Investigation Into The Establishment Of Operations Support Systems Permanent Performance Measures For Incumbent Local Exchange Telecommunications Companies." Florida Public Service Commission, Docket No. 000121-TP, volume 6, Cross examination of Dr. Robert Michael Bell, page 1103, lines 15 - 24.)

Q. WHAT METHOD DOES BELLSOUTH USE FOR BALANCING A PROPORTION MEASURE?

A. BellSouth's uses a concept called the "odds" ratio to set the alternative hypothesis for balancing a proportion measure.

Q. WHAT IS AN "ODDS" RATIO?

A. The "odds" ratio is what BellSouth has used when the information in the "cells" involves proportions, which I have been discussing, rather than "means." The "odds" methodology is relatively straightforward. First we need to define the odds of an event such as a missed installation occurring. Odds are the ratio of the probability of an event occurring to the probability that the event won't occur. So, if BellSouth "missed" 21.6 percent of the installations to their own customers, then the odds of a customer experiencing a "miss" is found by dividing the probability of a "miss," 0.216, by the probability of an "on-time" installation, 0.784 (= 1 – 0.216). This gives the odds of a "miss" as 0.276. In odds terminology, we might say that the odds of a BellSouth customer experiencing a "miss" are approximately 1 to 3.6.

The odds ratio for "missed" provisioning installations is the CLEC customer's odds of a "miss" divided by the BellSouth customer's odds of a "miss." When this odds ratio is one or less, BellSouth is delivering parity or better service to the CLEC's customers. When this odds ratio is greater than one, then BellSouth is not necessarily delivering parity service. Under a balancing approach, we need to determine an odds ratio greater than one to use for the balancing alternative hypothesis.

Q. IS THE ODDS RATIO EASY TO INTERPRET?

A. Not necessarily. Many people have trouble interpreting odds, and relating the value back to the probability of an event occurring. However, the interpretation in terms of odds is straightforward. If the odds ratio for "missed'

installations is set at 3, then we know that a CLEC customer's odds of a "miss" is three times greater than that of a BellSouth customer. We would still need a table, such as Dr. Bell's Table 2, to interpret the actual difference in the performance.

Q. CAN YOU PROVIDE US WITH SUCH A TABLE?

A. Certainly. Figure 1 below will help one interpret the actual difference between the BellSouth proportion and the CLEC proportion for a given "odds" ratio. The table shows the percentage of the time a CLEC customer will experience a miss by the BellSouth percentage "missed," for two values of the odds ratio: 2 and 3.

**Figure 1**
**CLEC Percentage of "Missed" Installations**
**By BST Percentage and**
**The Odds Ratio of the Alternative Hypothesis**

| BST PERCENTAGE MISSED | Odds Ratio 2 | 3 |
|---|---|---|
| 1 | 2 | 3 |
| 5 | 10 | 14 |
| 10 | 18 | 25 |
| 20 | 33 | 43 |

We see from the first row of this table that for an alternative hypothesis with an odds ratio of 3, the CLEC percentage of "missed" installations is about 3 percent when the BST percentage is 1 percent. However, the CLEC percentage is about 43 percent when the BST percentage is 20 percent. So when the BST percentage is close to 0, the CLEC percentage is about 3 times larger at the

balancing alternative hypothesis. As the BST percentage get larger, the ratio of the CLEC percentage to the BST percentage gets smaller; converging to 1 as the BST percentage approaches 100 percent.

Q. THIS SEEMS TO SUGGEST THAT IF BELLSOUTH HAS A MISS OF 20 PERCENT, THAT A MISS OF UP TO 43 PERCENT WOULD BE ACCEPTABLE FOR THE CLECS. IS THIS CORRECT?

A. No, that misses the point completely. With numbers like that, with a very small sample size, the methodology would show BellSouth out of parity almost 60 percent of the time and as the sample size approached a thousand transactions for BellSouth and only fifty for the CLEC, the probability that parity will not be concluded approaches 100 percent (see Figure 3 below). I realize this is not intuitive, and I will discuss it more below, but it would be a mistake to conclude that the odds ratio balancing test allows the CLECs to experience significantly worse performance than BellSouth without detecting a failure to provide parity on BellSouth's part. I would also note that the same holds true for Dr. Bell's calculations using the arcsine square root method where he shows a similar disparity. Once the sample size gets to the levels that I have just mentioned, the probability of finding a disparity at those levels approaches 100 percent.

Q. IF THE ODDS RATIO METHOD IS USED FOR DEFINING THE BALANCING CRITICAL VALUE, HOW DOES THAT EFFECT THE FORMULA THAT IS USED TO CALCULATE THE CRITICAL VALUE?

A. The balancing critical value for a proportion measure is based on a different formula than that of a mean measure when an odds ratio approach is used. The formula is more complicated than the mean measure formula, and it is given in Appendix C of the Louisiana "Statistician's Report."

Q. DR. BELL SUGGESTS IN HIS TESTIMONY THAT THERE IS A PROBLEM THAT CAN ARISE WHEN THE DELTA VALUE IS SET TOO LARGE. HE GOES ON TO SAY THAT PLACING A LOWER LIMIT ON THE SIZE OF THE CRITICAL VALUE COULD SOLVE THIS PROBLEM. PLEASE RESPOND.

A. In discussing large negative Z scores that do not trigger a test failure because the balancing critical value is larger (further from zero than the Z score), Dr. Bell states on page 40, lines 16-17, "Such an outcome would be justified only if one could be certain that delta has not been set too large." He goes on to say that he feels no floor (lower limit) is warranted if the "delta" he advocates, 0.25, is used.

From this statement, I infer that Dr. Bell understands that a balanced test has sufficient power to detect truly discriminatory performance on the part of BellSouth. However, this will only be true if "delta" is chosen so that it effectively defines the materiality threshold.

While I agree that "delta" needs to be chosen carefully, I am not convinced that

a "delta" of 0.25 is correct. The Louisiana Public Service Commission has ordered BellSouth to use a "delta" of 1. The Georgia Public Service Commission has ordered that a "delta" of 0.5 be used. In both situations, there will be periodic reviews of the effectiveness of the methodology. I assume that if these commissions find that "delta" was set too large, they will lower the value.

To me, placing a floor on the balancing critical value is trying to solve a problem that does not exist. Once "delta" is chosen, it should be understood that BellSouth would be found out of parity any time the observed difference in mean performance is larger than one-half "delta" standard deviations. This creates a test that has a lot of power to detect disparities beyond one-half "delta," but almost no power to detect disparities less than one-half "delta." If one considers this when choosing "delta" then there should be no reason to protect against a situation where "delta" is set too large.

Q. YOU SAID THAT A TEST BASED ON BALANCING HAS A LOT OF POWER TO DETECT DISPARITIES BEYOUND ONE-HALF DELTA. WOULD YOU PROVIDE US AN EXAMPLE OF THIS?

A. Yes. Figure 2 shows the probability that a mean measure statistical test will detect a difference in the mean performance of BellSouth and a CLEC when the balancing alternative hypothesis uses a "delta" of 1. To calculate these we assume that the true disparity is 0, 0.2, 0.45, etc. For the purpose of this example I am defining the "true disparity" as the numbers indicated across the

top of the chart. This is not an observable figure; I am assuming the disparity to exist to illustrate what I am talking about. If we have used a delta of 1, this chart would tell us that any "true discrepancy" below 0.5 is immaterial and any "true discrepancy" above 0.5 is material. The chart shows the probability of detecting this condition. Using an example from the chart, assume a very small sample size, which is always going to be problematic. In the first line, even if the "true disparity" was zero, that is there was no disparity, the statistical analysis is going to show that there is disparity 32 percent of the time. On the other end of the scale, at 1, the analysis is only going to show a material difference 68 percent of the time, when we know that the disparity actually exists and is material. These are essentially examples of Type 1 and Type II errors, where the Type II error at the 1 disparity level is 32 percent (the complement of the probability of detection). Importantly, as the sample size increases, the analysis rapidly approaches an accuracy level of 100 percent, meaning that the Type I and Type II errors are essentially eliminated.

**Figure 2: The Probability of Detecting Disparity**
**Mean Measure Test with Delta = 1**

| BST Sample Size | CLEC Sample Size | Balancing Critical Value | True Disparity Level | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 0.2 | 0.45 | 0.5 | .55 | 0.8 | 1 |
| 10 | 1 | -0.477 | 0.317 | 0.387 | 0.481 | 0.5 | 0.519 | 0.613 | 0.683 |
| 100 | 5 | -1.091 | 0.138 | 0.256 | 0.457 | 0.5 | 0.543 | 0.744 | 0.862 |
| 1000 | 50 | -3.45 | 0 | 0.019 | 0.365 | 0.5 | 0.635 | 0.981 | 1 |
| 12000 | 800 | -13.693 | 0 | 0 | 0.085 | 0.5 | 0.915 | 1 | 1 |
| 100000 | 2500 | -24.693 | 0 | 0 | 0.007 | 0.5 | 0.993 | 1 | 1 |

Q. IT SEEMS THEN THAT A MEAN MEASURE TEST BASED ON A BALANCING METHODOLOGY DOES MAKE IT POSSIBLE TO DETECT

DISCRIMINATION AS LONG AS THE TRUE DISPARITY IS BEYOND THE MATERIALITY THRESHOLD. IS THAT TRUE?

A. Yes, a mean measure test based on balancing and large sample sizes has a high likelihood of detecting disparity beyond the one-half "delta" materiality threshold, but a low probability of detecting disparity that falls under the threshold.

Q. ISN'T IT TRUE THAT THESE CONDITIONS ARE THE SAME ONES THAT LEAD TO BALANCING CRITICAL VALUES THAT ARE FURTHER FROM ZERO THAN THOSE THAT ARE CONVENTIONALLY USED?

A. Yes. Large sample sizes lead to critical values that are further from zero than those that are used in many applications. Such critical values, in turn, lead to small significance levels. But, as I have shown, those small significance levels (which are the probabilities corresponding to a true disparity of 0 in Figure 2) do not imply that BellSouth will get away with any amount of discrimination. Those levels of disparity that are lower than the penalty payment (or materiality) threshold of one-half delta will not be considered discriminatory. However, levels of disparity beyond the materiality threshold will be detected as discriminatory with a high likelihood.

Q. IS THE SAME THING TRUE FOR PROPORTION MEASURES?

A. A similar statement can be made for a proportion measure test. When using an odds ratio approach to balancing, the materiality threshold is not one-half of the odds ratio used in the balancing alternative hypothesis, but the threshold is at a point close to this. Figure 3 below illustrates this by showing the probability that the testing procedure will determine disparity (reject the null hypothesis), for a range of disparity levels and BST/CLEC sample sizes when the BellSouth proportion of missed installations is 0.20 and balancing is done for the alternative hypothesis with an odds ratio of 3.

Notice that for a balancing alternative with odds ratio of 3 (BST proportion of 0.20 and CLEC proportion of 0.43), there is a significant probability of determining disparity for odds ratio levels less than 3. For example, with a CLEC proportion of misses of 0.30 there is at least a 50% chance, regardless of sample size, that disparity will be determined and a remedy paid. Here we have an odds ratio of 1.75, much less than the balancing alternative of 3.

**Figure 3: The Probability Of Determining Disparity When the BellSouth Proportion of Missed Installations is 0.20 and the Balancing Critical Value is Determined at an Odds Ratio of 3**

| **Number of Transactions** | | **Level of Disparity in Terms of Odds Ratio** *Level of Disparity in Terms of CLEC Proportion* | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **1*** | **1.25** | **1.75** | **2** | **2.25** | **2.75** | **3**** |
| BST | CLEC | *0.20* | *0.24* | *0.30* | *0.33* | *0.36* | *0.41* | *0.43* |
| 10 | 1 | 0.4110 | 0.4440 | 0.5000 | 0.5220 | 0.5410 | 0.5750 | 0.5890 |
| 100 | 5 | 0.2920 | 0.3730 | 0.5040 | 0.5570 | 0.6030 | 0.6790 | 0.7080 |
| 1000 | 50 | 0.0410 | 0.1530 | 0.5130 | 0.6750 | 0.7960 | 0.9300 | 0.9590 |
| 12000 | 800 | 0.0000 | 0.0000 | 0.5520 | 0.9640 | 0.9990 | 1.0000 | 1.0000 |
| 100000 | 2500 | 0.0000 | 0.0000 | 0.5930 | 0.9990 | 1.0000 | 1.0000 | 1.0000 |

* An odds ratio of one assumes that there is parity. Thus, the probability of determining disparity in this situation is the probability of a Type I error.

** The probability of determining disparity increases as the level of disparity goes beyond an odds ration of three.

Q PLEASE RECAP YOUR POINT REGARDING DR. BELL'S TESTIMONY THAT YOU HAVE BEEN DISCUSSING.

A. Dr. Bell seems to feel that if "delta" is chosen too large then material disparities will not be discovered, particularly with large sample sizes. To protect against this, he feels that a floor on the balancing critical value may be needed. I do not see that this solves any problems. If "delta" is set too large, then it is too large regardless of the sample sizes that are employed in the test. If one understands how a statistical test that employs balancing works, then there should be no need to place an artificial floor on the balancing critical value.

Q. WOULD YOU DISCUSS MATERIALITY AGAIN IN THE CONTEXT THAT WE ARE USING THE TERM IN THIS PROCEEDING?

A. Certainly. Recall from my direct testimony that as long as the average time taken to provide the relevant service to a CLEC does not exceed the BellSouth mean plus one-half "delta" times the BellSouth standard deviation, then the apparent difference in mean service times would not be material. That is, we would not conclude that BellSouth is providing discriminatory service. To state this another way, one-half delta, where delta is the parameter that defines the alternative hypothesis for balancing, is a materiality threshold for the disparity in the service system when a balancing method is used for a mean measure test.

Q. IN HIS TESTIMONY, DR. BELL ARGUES THAT THE LOUISIANA REPORT IMPLICITLY DEFINES MATERIALITY AS THE VALUE OF DELTA. DO YOU AGREE?

A. No. "Delta" is a factor that is used to identify whether a meaningful difference exists between the BellSouth and CLEC performance. As an author of the Louisiana "Statistician's Report," I do not believe it was our intention to make materiality synonymous with the value of "delta." Materiality is directly related to the parameter "delta." As I explain above, penalty payments apply when the observed disparity is more than one-half "delta." So when one chooses "delta" for the alternative hypothesis, then you automatically set the materiality threshold to be one-half "delta."

If "delta" were actually the materiality threshold, then both the SEEM and AT&T performance plans are calling for payments to be made on service differences that are immaterial (differences in average performance between one-half "delta" and "delta" standard deviations). Consider Dr. Bell's example where the average BellSouth time to complete an order is 5 days with a 5-day standard deviation. If "delta" is set at 0.25, as in Dr. Bell's example, then the Type II error probability used for balancing is evaluated assuming the CLEC average completion time is 6.25 days. This means that once the CLEC average completion time goes beyond 5.625 days (a 15 hour difference between the BellSouth and CLEC average service times), then BellSouth would pay a penalty. If in fact the commission determined that 1.25 days (or 30 hours) constitutes a material difference in the average service times, then why should

a penalty be paid for immaterial differences between 15 and 30 hours?

If you carefully go through the penalty plan proposed by AT&T in Ms. Bursh's testimony, you find that an observed disparity of "delta" is labeled an intermediate failure for a Tier I comparison, and it is labeled market impacting for a Tier II comparison. Furthermore, Tier I penalties for observed disparities between one-half "delta" and "delta" range from $2,500 to $8,125. Tier II penalties range from $5,000 (when the observed disparity is five-sixths "delta" and the market penetration factor is one) up to $81,250 (when the observed disparity is "delta" and the market penetration factor is ten). These classifications and penalty amounts do not suggest that AT&T agrees with Dr. Bell that disparities less than "delta" are immaterial.

Q. DR. BELL EXPRESSES CONCERN REGARDING THE SEEM REMEDY CALCULATION. WOULD YOU PLEASE EXPLAIN THE CONCEPT BEHIND THE VOLUME PROPORTION CALCULATION THAT IS USED TO DETERMINE THE PENALTY PAYMENT ADVOCATED BY BELLSOUTH?

A. Certainly. It is my understanding that BellSouth wants to have a penalty plan based on the number of CLEC transactions that, if changed for the better, would change a Truncated Z test failure into a pass. That is, under BellSouth's plan, a calculation is made of the number of transactions that would have had to be accomplished more quickly (if the time interval was the relevant measure) in order to avoid having a failure. The penalty is then determined by

multiplying that number of transactions times the appropriate penalty amount. This concept is used in Southwestern Bell's Texas penalty plan. The problem for BellSouth is that the Truncated Z statistic is much more complex than the statistics that Southwestern Bell uses because the Truncated Z has some desirable properties that the Texas plan's statistics do not have. This makes it very difficult to directly observe or calculate the number of transactions in any particular situation that would have to be changed to turn a failure into a pass. That being the case, BellSouth developed a surrogate model for calculating the number of transactions subject to remedy.

Q. WOULD YOU EXPLAIN HOW BELLSOUTH'S SURROGATE MODEL WORKS?

A. Yes. When BellSouth fails a test, that is the Truncated Z statistic is found to be less than the critical value (further from zero in the negative direction), the calculation of the penalty amount proceeds as follows:

1. Determine the "impacted volume," that is the number of CLEC transactions in cells with negative z-values that are potentially subjected to penalties. These are transactions that, if changed for the better, would increase a Truncated Z value in the positive direction.

2. Calculate the "parity gap." Subtract the Truncated Z statistic from the balancing critical value. For example, if the Truncated Z statistic for a particular submeasure is calculated to be –3, and the balancing critical

value is –2, then there is a test failure. The "parity gap" is the distance between –2 and –3, which is 1 unit. If the Truncated Z statistic turned out to be –4 instead, then the "parity gap" would be 2 since the distance between –4 and –2 is two units.

3. Calculate the volume proportion. Divide the "parity gap" by 4. If the "parity gap" is larger than 4, than the volume proportion is capped at 100%. Otherwise, the "volume proportion" will be between 0 and 100%. For example, when the "parity gap" is 1, then the "volume proportion" is 1 divided by 4, or 25%. If the "parity gap" is 2, then the "volume proportion" is 50%. The volume proportion keeps growing, as the gap gets larger. Once the gap reaches 4 units, the volume proportion reaches 100%. Since you could not change any more impacted transaction for the better, we stop here, and set the volume proportion to 100% for all gaps larger than 4 units.

4. Calculate "effected volume" by multiplying the "impacted volume" by the "volume proportion." This is the number of transactions for which a penalty will be paid. The "effected volume" is the surrogate model's estimate for the number of transactions to change in order to get the Truncated Z value equal to the balancing critical value. For example, if the "impacted volume" is 100 CLEC transactions, and the "volume proportion" is 50%, then the "effected volume" is 50 transactions.

5. Calculate the final penalty payment. Multiply the "effected volume" by the per transaction fee of the submeasure class (given by BellSouth witness Mr. Varner in Exhibit No. AJV-2).

Q. WHAT DO YOU MEAN BY TRANSACTIONS THAT ARE "CHANGED FOR THE BETTER?"

A. The term "change for the better" can mean different things for different types of measures. For a proportion measure, such as missed installations, "poor" transactions are those that are not completed on time (they are "missed"). Changing a "poor" transaction for the better would amount to changing a "missed" transaction into a "non-miss," or completed on time. So for a proportion measure we would want to start to define the impacted volume with all "missed" CLEC transactions. If we are dealing with the rate measure customer trouble report rate, a "poor" transaction is any trouble, and if the trouble had never occurred, that would be "better." So for this rate measure, defining the impacted volume starts with all CLEC troubles. For a metric that is measuring the average duration of a service, such as order completion interval, a change for the better means that the service time is made shorter. This could be any CLEC service order, so for mean measures we would want to start to define the impacted volume with all CLEC transactions.

Q. WHY ARE TRANSACTIONS FROM CELLS WITH NEGATIVE Z-VALUES THE ONLY ONES USED TO DEFINE IMPACTED VOLUME?

A. Only transactions from negative cells that are "changed for the better" can increase the Truncated Z value in the positive direction. If we used all CLEC transactions that could be changed for the better to define the impacted volume, we would be including transactions that will not increase the Truncated Z value even if the transactions are "changed for the better." Recall, that in order to compute the Truncated Z statistic, we disaggregate the data into many like-to-like cells, and compute cell z-values. In calculating the value of the Truncated Z, positive cell z-values are set to zero. These cells do contribute to the final value of the Truncated Z, but changing "poor" transactions to "better" values in such cells, will not change how much these cells contribute to the Truncated Z. This is because these cells already exhibit service that is favorable to the CLEC. Changing transactions for the better will still result in a positive cell z-value that is still set to zero for calculating the Truncated Z. Our goal is to increase the value of the Truncated Z result by changing transactions for the better. This will not happen for cells with positive z-values. Thus, only CLEC transactions in cells with negative z-values can be changed for the better, and improve the value of the Truncated Z score. So, the impacted volume is the number of CLEC transactions in "negative" cells that can be changed for the better.

Q. IS THERE A WAY TO CALCULATE THE NUMBER OF TRANSACTIONS WITHOUT THE USE OF A SURROGATE MODEL?

A. Yes, if you have a proportion or rate measure. It turns out that a mathematical solution, called a linear program, can be used to find the maximum number of

"missed" transactions in negatively effected cells that, if changed to "non-misses," will shrink the gap between the balancing critical value and the Truncated Z statistic to zero. Linear programs can be very time consuming to solve, especially if the Truncated Z test involves a large number of cells. Thus, one may still want to use a surrogate model to be able to calculate the transaction count. Even if a linear program is used for proportion and rate measures, it will not work for a mean measure. Using the same surrogate model for all measure types seems like a reasonable thing to do.

Q. HAVE YOU EVER USED THE LINEAR PROGRAMMING METHOD TO CALCULATE THE NUMBER OF TRANSACTIONS FOR THE PENALTY CALCUATION AND COMPARED THAT TO THE RESULTS OBTAINED BY USING BELLSOUTH'S SURROGATE METHOD?

A. Yes. The Ernst & Young statistical team examined Louisiana data from November 1999 for the percent missed installations performance measurement. We were able to solve the linear program for 39 Tier I and Tier II tests in which BellSouth failed a parity test.

Q. WHAT WERE THE RESULTS OF YOUR ANALYSIS?

A. For each of the 39 tests, BellSouth's SEEM plan (the surrogate model) calculated the same number, or a higher number, of transactions for penalty payment than the linear program solution. In fact, the linear program calculated a total of 352 transactions for penalties across the 39 tests. The

SEEM plan, on the other hand, calculated a total of 814 transactions for penalties. This means that the SEEM plan was calculating more transactions to "change for the better" than needed for the purpose of shrinking the gap between the Truncated Z statistic and the balancing critical value.

Q. WOULD YOU SUMMARIZE THE KEY ISSUES THAT THE COMMISSION NEEDS TO CONSIDER IN ADOPTING THE METHODOLOGY YOU ARE RECOMMENDING?

A. Yes. In order to carry out the Truncated Z with Error Probability Balancing, the Commission needs to evaluate two key aspects of any proposed plan: 1) the level of aggregation at which parity decisions will be made, and 2) the "delta" value used to determine the balancing critical value. Neither of these input parameters is something that should be decided upon solely by statisticians. Input from subject matter experts is needed.

While I do not have a specific recommendation for the "delta" value, I believe that it should be understood that the impact of any choice of "delta" is that BellSouth will pay penalties when the observed difference in the means of the ILEC and CLEC is greater that one-half "delta" standard deviation, as described in the examples I have already given.

Q. DOES THIS CONCLUDE YOUR REBUTTAL TESTIMONY?

A. Yes.