

## Use of the Chi-square Test to Determine Significance of Cumulative Antibiogram Data

<sup>1</sup>Rocco J. Perla and <sup>2</sup>James Carifio

<sup>1</sup>HealthAlliance Hospital, Leominster, Massachusetts, USA

<sup>2</sup>University of Massachusetts, Lowell, USA

---

**Abstract:** An important function of a hospital's Infectious Disease and Pharmacy programs is to review and compare the most recent antibiogram with that of the previous year to determine if significant changes in antibiotic susceptibility results are noted and to communicate this information and its consequences to the medical staff. However, there are currently no formal analytical (decision-making) models in use to determine if the rate of resistance to an antibiotic from one year to the next has significantly changed more or less than one would expect due to sampling error and test reliability. The purpose of this article, therefore, is to demonstrate the utility of using a well-established and simple nonparametric statistical technique (chi-square) for analyzing annual variations in cumulative antibiogram data and to determine whether such variations are significantly different from chance and to what degree. The chi-square model outlined here is a simple, practical, quick, low burden and easy to understand and execute approach that greatly improves the analysis of antibiogram data and decision-making by practitioners. More work and research is needed to develop additional inferential statistical methods and models that can be applied to antibiogram data.

**Key words:** Antibiogram, chi-square, decision-making, nonparametric, statistic

---

### INTRODUCTION

The Clinical and Laboratory Standards Institute (CLSI, formerly the National Committee for Clinical Laboratory Standards) defines an antibiogram as an overall profile of antimicrobial susceptibility results of a microbial species to a battery of antimicrobial agents<sup>[1]</sup>. Antibiograms have long been used as an epidemiological tool to characterize the susceptibility patterns and profiles of bacterial species over time in clinical settings and they are also believed to play an important role as a guide to empiric antimicrobial therapy<sup>[1]</sup>. Antibiograms are most often presented in the form of large 2 x 2 tables that compare different organism-antimicrobial agent susceptibility combinations in a one-to-one correspondence. The data in standard antibiogram tables lists the percentage of isolates for a single bacterial species that are susceptible to an array of different antibiotics. It is recommended that antibiograms be created on an annual basis in order to compare susceptibility results for a bacterial species versus specific antibiotics over time<sup>[1]</sup>. Indeed, an important function of a hospital's Infectious Disease and Pharmacy programs is to review and compare the most recent antibiogram data with that of the previous year to determine if significant changes in antibiotic susceptibility results are noted. It is expected that "significant" changes in a hospital antibiogram will be communicated to physicians who in turn will consider these changes when prescribing antibiotics empirically. However, the antibiogram review process has been, for

the most part, an informal and somewhat subjective task requiring numerous comparisons of antibiotics and bacterial species. There are currently no formal analytical (decision-making) models in use to determine if the rate of resistance to an antibiotic from one year to the next has significantly changed more or less than one would expect due to sampling error and test reliability. For example, what does it mean statistically, logically and interpretively to say that the rate of ampicillin resistance in *Escherichia coli* in Hospital A went from 36% in 2004 to 41% in 2005? This question becomes all the more important considering that the different methods used to do routine susceptibility testing in laboratories, although reliable, are somewhat variable and a perfect test-re-test correlation cannot be assumed<sup>[2-4]</sup>. The purpose of this article, therefore, is to demonstrate the utility of using a well-established and simple nonparametric statistical technique for analyzing annual variations in cumulative antibiogram data and to determine whether such variations are significantly different from chance and to what degree.

**Chi-square and independent samples:** Chi-square ( $\chi^2$ ) is one of the most widely used statistical tests for nominal (categorical) data and has been applied to a wide range of issues and problems where frequency data is involved<sup>[5]</sup>. One of the key requirements of the chi square test is that the data categories are independent and mutually exclusive. Antibiogram data is nominal data in that a particular bacterial species (e.g., *E. coli*, n = 1000) may include two sub-

populations relative to any particular antibiotic tested against the bacteria (i.e., a resistant sub-population and a susceptible sub-population). Antibigram data is also independent and mutually exclusive. For example, if 36% of *E. coli* in a given year are resistant to ampicillin, then the remaining 64% are susceptible to ampicillin. In some cases, a bacterial isolate may be neither resistant nor susceptible to an antibiotic, but demonstrate intermediate level susceptibility. In most cases, intermediate isolates are included in the resistant category, as such isolates are non-susceptible. The recommendations established for the development and analysis of cumulative antibiograms by the CLSI dictate (1) that a single patient may contribute only a single unique bacterial species in a one-year period to the data pool and (2) that only clinically relevant cultures (i.e., excluding surveillance or screening cultures) be incorporated into the analysis and category designations, thereby eliminating samples that may artificially skew the susceptibility data<sup>[1]</sup>.

**The  $\chi^2$  formula:** An example of how the chi-square statistic is used in cumulative antibiogram analysis is provided below. We want to know if the observed resistant and susceptible rates for a particular bacterial species relative to a particular antibiotic during one year are significantly different from the rates we expect to observe. In order to make this determination, we use the chi-square test. The chi-square value is calculated using the following standard equation:

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

The  $\chi^2$  value above is derived from the sum of the observed values minus the expected values squared ( $fo - fe$ )<sup>2</sup> divided by the expected value ( $fe$ ). Clearly, as the  $\chi^2$  equation demonstrates, if the observed values are equal to the expected values, the  $\chi^2$  value is zero indicating no difference between what we observed and what we expected to observe based on the  $\chi^2$  distribution probabilities.

**Expected values:** In the  $\chi^2$  analysis, the expected values can generally be derived through one of three ways: through chance (probability), through an a priori theory or hypothesis, or through existing data and research. The latter model is often referred to as the “empirical” model and is the model used here for generating the expected antibiogram values. For over a decade, most clinical microbiology laboratories have created and disseminated antibiogram data and tables, resulting in an extensive “local” data repository. Due to variations in different hospital settings (such as level of acuity and patient populations), antibiogram data tends to vary between hospitals making local historical data the best estimate of expected antibiogram values for a

particular hospital. Because antibiogram data generally tend not to vary over short time periods (e.g., 2-3 years vs. 5-10 years), especially if a moderate degree of resistance is already established<sup>[6]</sup>, the average antibiogram data for a hospital during the three year period prior to the year in question is a reasonable choice to be used as the expected value. This three year average is the least burdensome and most easily obtained empirical expectancy estimate and it is similar to the moving three year averages that are often used in many different types of trend analyses in several different areas for data that is somewhat chaotic or noisy due to several uncontrolled random factors that net out to a random effect in 99.9% of the cases<sup>[7,8]</sup>. One may average a larger number of prior data points to obtain the empirical expectancy estimate, but ones needs to be careful about the number of points one uses to create the empirical expectancy estimates, as averaging out many data points may both blunt and disguise rapidly emerging “local” changes as opposed to large “global” changes and major parameter shifts. The goal here, however, is to detect statistically significant shifts in rates rapidly and reliably and as they are emerging in the categories of interest, by creating an empirical criteria for objectively rather than subjectively evaluating and making decisions about the percentages.

It should be noted that one of the assumptions of  $\chi^2$  is that no expected category should be less than one<sup>[5]</sup>. This is an important point to consider regarding antibiogram data because if we do not expect to see resistance to an antibiotic by a specific bacterial species *at all* in an institution (such as vancomycin resistance in *Staphylococcus species*), then the presence of resistance *to any degree* is significant (clinically and epidemiologically) regardless of the sample size, metric, test statistic or level of significance used. The same rationale would hold for bacterial species and antibiotics where we would expect virtually complete resistance (such as ampicillin resistance in *Klebsiella species*).

**Sample size:** One of the important assumptions of  $\chi^2$  is a sufficiently large sample. Applying  $\chi^2$  to small samples increases the risk of Type II errors to an unacceptable level<sup>[5]</sup>. Sample size is not usually an issue with antibiogram data as the sample sizes are most often large (greater than 100), although infrequently encountered bacteria will obviously have lower testing frequencies. Nevertheless, the CLSI M39-A2 document now recommends that antibiogram analysis only be done on bacterial species with a frequency of 30 or greater. This criterion is reasonable to apply to the  $\chi^2$  analysis of antibiogram data discussed here and corrective adjustment procedures may be used for data from samples of less than 30 observations.

**Example:** This section of the article demonstrates the process by which antibiogram data can be translated into a  $\chi^2$  statistic that can be used to determine whether the average susceptibility data derived from a hospital antibiogram over the last three year period for individual bacteria-antibiotic comparisons (*fe*) is different from the most recent antibiogram data (*fo*). The example used here represents actual antibiogram data analyzed by the authors from a community hospital setting following the M39-A2 guidelines for antibiogram preparation. For illustration purposes, only a single bacterial species (*Pseudomonas aeruginosa*)/single antibiotic (ceftazidime) comparison is made. This selection was made because *P. aeruginosa* is a critically important human pathogen and ceftazidime is a drug considered to be highly effective against this species<sup>[9]</sup>. Consequently, annual variations in ceftazidime susceptibility to *P. aeruginosa* need to be understood in a non-arbitrary manner.

The problem we are faced with is as follows and is similar to situations that often arise in the area of clinical microbiology and infectious disease epidemiology for many different bacteria-antibiotic comparisons. Specifically, in 2004, *P. aeruginosa* was susceptible to ceftazidime 83% of the time—an 11% decrease in ceftazidime susceptibility from the previous year, which was 94%. We also know that the rates of ceftazidime susceptible *P. aeruginosa* from 2001 through 2003 were 89% (n=118), 91% (n=128) and 94% (n=127), respectively. Intuitively it appears (1) that ceftazidime resistance in *P. aeruginosa* from 2001 through 2003 was stable if not slightly declining and (2) that a significant reduction in ceftazidime susceptible *P. aeruginosa* occurred from 2003 to 2004. But just how significant is this reduction and could this reduction be sampling error or/and chance variation and not a real or rapid negative shift in susceptibility? These are questions the  $\chi^2$  analysis answers.

Table 1 shows the values used in the  $\chi^2$  calculation. The null hypothesis in this situation is that the 2004 *P. aeruginosa* ceftazidime susceptibility results do not differ significantly from the average susceptibility results from the prior three years. Thus, the expected rate of ceftazidime resistance is 9% of the 121 *P. aeruginosa* isolates, or  $0.09 \times 121 = 11$  cases. The expected rate of ceftazidime susceptible *P. aeruginosa* is 91% of 121, or  $0.91 \times 121 = 110$  cases. Using these counts, the  $\chi^2$  value is calculated to determine if the susceptibility rates between the expected and observed data is significantly different. Because one of the cells in Table 1 (Expected cases/Resistant) is small and because there is only one degree of freedom, the Yates correction (although controversial) is applied to this case to reduce the chance of artificially increasing  $\chi^2$  thereby making it more difficult to establish significance and thus reducing Type I error (accepting a false hypothesis) in this particular type of decision-

making context and situation where one may need to be more statistically “conservative” than in other situations (see text below).

Adding the quotients in Table 1 (values in the last row) gives us a  $\chi^2$  value of 9.02 with one degree of freedom. This value is significant at the .01 level ( $p = 0.0026$ ) and statistically confirms suspicion that the reduced *P. aeruginosa* susceptibility to ceftazidime in 2004 was significant compared to our expected value, which is the average susceptibility rates during the last three years. In statistical terms, this  $\chi^2$  value leads us to reject the null hypothesis, which stated that there was no significant difference between the observed and expected frequencies. This  $\chi^2$  value provides the *formal* warrant to explore issues that may be leading to decreased ceftazidime susceptibility for *P. aeruginosa*, such as the make up of the hospital formulary, prescribing practices of physicians and susceptibility trends at other local hospitals.

In the example in Table 1, both intuition and the  $\chi^2$  statistic are consistent with one another. That is, we intuitively expected the difference in susceptibility rates to be significant and our statistic confirmed this suspicion. However, most antibiogram comparisons are not so intuitive leaving us with less confidence in our informal judgments and this situation and case is where statistics is most helpful. We can easily imagine a situation where the observed values described above (83% susceptible and 17% resistant) vary. In Table 2, the susceptibility values for *P. aeruginosa* and ceftazidime are modified by 2 *percentage points* to demonstrate how the  $\chi^2$  statistic can assist in making antibiogram decisions when the data is not so intuitive or obvious.

Table 2 shows that as the percent of ceftazidime susceptible *P. aeruginosa* isolates increases by 2%, the  $\chi^2$  values decrease (keeping *fe* constant). The observed data analyzed in Table 1, which is the first entry in Table 2 (17%R and 83%S) is significantly different from the expected values at both the .05 and .01 level. Obviously, these results are statistically significant. However, if we increase the susceptibility percentage from 83% to 85%, it becomes more difficult to *informally* determine if this value is significantly different from the historical data at the hospital (i.e., the prior three year period). This difficulty is represented in Table 2 where an 85% susceptible rate is significant at the .05 level but not at the .01 level. How should this result be interpreted?

When dealing with antibiotics that are critically important in treating infections caused by certain bacterial species (as is the case for ceftazidime in treating *P. aeruginosa* infections), it seems appropriate to set alpha at .05 (vs. .01) and to tolerate a higher risk of sampling error when the decision-making preference is to accept and act on a “slight or close” variation between the expected and observed antibiogram values. For bacterial species-antibiotic comparisons that are deemed to be less critical at any point in time, it appears

Table 1: Values used in the  $\chi^2$  calculation.<sup>a</sup>

	Resistant	Susceptible	Total
Observed cases ( <i>fo</i> ) (2004)	21 (17%)	100 (83%)	121
Expected cases ( <i>fe</i> ) <sup>b</sup> (prior 3 year average)	11 (9%)	110 (91%)	121
<i>fo-fe</i>	9.5	-9.5	0
( <i>fo-fe</i> ) <sup>2</sup>	90.25	90.25	
( <i>fo-fe</i> ) <sup>2</sup> / <i>fe</i>	8.2	0.82	$\chi^2_{.01(1)} = 9.02$

<sup>a</sup>Yates correction applied.

<sup>b</sup>Expected values based on the average of the three years prior to 2004 (2001, n = 118; 2002 n = 128 and 2003, n = 127).

Table 2: Modified observed *Pseudomonas/ceftazidime* susceptibility data and associated  $\chi^2$  values keeping *fe* constant at 9% resistant and 91% susceptible with 1 degree of freedom (n = 121). Yates correction applied.

As percentage (%R/%S)	Observed data As actual counts (R/S)	$\chi^2$	Significant at alpha		
			.10	.05	.01
17/83	21/100	9.02	Yes	Yes	Yes
15/85 <sup>a</sup>	18/103	4.22	Yes	Yes	No
13/87 <sup>a</sup>	16/105	2.02	No	No	No
11/89 <sup>a</sup>	13/108	0.22	No	No	No

<sup>a</sup>Hypothetical data for the basis of graded comparison.

R = Resistant; S = Susceptible

reasonable to set alpha at .01 when the intent is not to accept and act on a slight or close variation between the expected and observed antibiogram values. The rationale for setting a more stringent alpha level (.01) for non-critical (or less important) antibiotic-bacterial species comparisons is that we want to be more confident that when we take action and expend limited healthcare resources on less important comparisons there is clearly good reason to do so and that there is a lower chance that we are acting on a variation based on sampling error.

Following this scheme and rationale, individuals responsible for determining the significance of annual variations in cumulative antibiograms would be responsible for determining which antibiotic and bacterial species comparisons are most significant and to set alpha accordingly. It would also seem helpful from a decision-making standpoint to code (perhaps by color) the squares in the antibiogram table that represent values that are significant at the .01 and .05 level and use this information to quickly and easily gauge the degree of variation from expectation in the antibiogram and to decide which data points require follow up. Arrows could be added to the squares to show the direction of the variation (i.e., arrow up for increased resistance and arrow down for decreased resistance), as bidirectional changes in antibiogram data can and do occur. Using a color and arrow scheme allows anyone looking at the data to quickly gauge both the direction and magnitude of the antibiogram variations. Although bidirectional shifts in antibiotic susceptibilities occur, the trend is clearly and alarmingly moving toward increased levels of resistance—an unavoidable consequence of bacterial evolution which is extremely complex and difficult to predict<sup>[10]</sup>.

It is important to point out that if the expected values used in the calculations in Table 1 and 2 were derived from the random probability model rather than the observation model, the  $\chi^2$  values would be much larger for the probability model and that this would represent false and misleading findings. Indeed, using the probability model (50%R and 50%S) for the expected values to analyze the data in the last column in Table 2 (11%R and 89%S) would generate a  $\chi^2$  value of 73 (a highly significant value with essentially a zero probability of occurring by chance), instead of a  $\chi^2$  value of 0.22 using the empirical three year average model, which is not significant at the .10 level (p = 0.64). Of course, for antibiotic and bacterial species comparisons that do approach a 50% susceptible rate, there would be less of a discrepancy between the probability model and the empirical model. However, antibiogram susceptibility results approaching 50% are not common, which is further justification for using the empirical model. Similarly, if only the prior year's data was used as the expectancy estimate (i.e., 2003, 6%R and 94%S) and thus the comparison point for 2004, the resulting  $\chi^2$  value of 56 would be highly significant with essentially a zero probability of occurring by chance. Again, because the phenomena of antibiotic resistance is highly complex, dynamic and somewhat chaotic<sup>[10]</sup> and because susceptibility test results are subject to variation within and between different testing systems<sup>[2-4]</sup>, comparisons using a single point expectancy estimate (e.g., the previous year's data) are likely to contain more sampling error than the most recent three year estimate, which is more "insulated" from unknown and random factors that may

sporadically impact the data (aside from the sources of variation that are well described and known).

## DISCUSSION

The  $\chi^2$  approach used here is a formal, systematic and non-capricious method to determine whether annual antibiogram data is significantly different from recently compiled data (i.e., the three previous years) as well as the magnitude of that difference. This model, like most statistical models, is most helpful in making decisions where large amounts of data need to be processed and summarized for the purposes of making important decisions. The type of analysis outlined here may be applied to the antibiogram as a whole or to specific susceptibility data where disagreement or lack of consensus regarding the significance of results is encountered. It would be simple and rather easy to create the  $\chi^2$  function used here in a software program such as Excel and to quickly analyze all antibiogram data for a hospital. As mentioned above, it would also seem helpful from a decision-making standpoint to code (perhaps by color) the squares in the antibiogram table that represent values that are significant at the .01 and .05 level and use this information to decide which data points require follow up. Arrows could be added to the squares to show the direction of the change. Using a color and arrow scheme allows anyone looking at the data to quickly gauge both the direction and magnitude of the antibiogram variations.

There are certainly other more sophisticated types of statistical analyses and models that could be applied to antibiogram data such as a z-test for comparing two independent proportions, logit analysis and Cramer's V to describe linear associations and trends for designs that are greater than the above standard 2x2 design. It should also be pointed out and stressed that nonparametric tests (such as  $\chi^2$ ) lack statistical sensitivity as compared to parametric tests and that the results of nonparametric tests need to be much larger to reach significance. Parametric tests are, therefore, much more powerful, much better at detecting weak effects and much better at giving accurate effect sizes or magnitudes of difference. For example, the counts in the third row in Table 2 (16 resistant and 105 susceptible) produced a  $\chi^2$  value of 2.02 which was not significant at the .1 level, however analyzing these same values using Fischer's z-test for independent proportions<sup>[11]</sup> produced a z score of 1.5, which is significant at the .1 level and very close to reaching significance at the .05 level ( $p = 0.06$ ).

Although parametric tests are more sophisticated and sensitive to weak effects, it is quite reasonable to use the easier and less sensitive (nonparametric)  $\chi^2$  model as outlined here to ensure that the differences observed between actual and expected antibiogram values are large enough to justify the time, cost and effort of following up on such changes, many of which

may represent the somewhat chaotic and unpredictable nature and fluctuations of antibiotic susceptibility over time. In instances where greater sensitivity in detecting changes in antibiotic susceptibility is desired, Fischer's z-test for independent proportions is a more logical metric than the less sensitive  $\chi^2$  test. The most effective antibiogram analysis model may actually involve the use of the nonparametric  $\chi^2$  test and Fischer's z test for independent proportions. Fischer's z test can detect early and subtle variations in susceptibility (and put one on notice), while the less sensitive  $\chi^2$  test can be used to determine if and when it is time to act. In this sense and in this context, the z test acts as an early warning system that gives one time to prepare for the situation and get a response in place and the  $\chi^2$  results signal that it is time to make a decision about whether or not to respond.

Regardless of the statistical model or models used to analyze antibiogram data, the decision to use a particular test, or series of tests and to establish alpha levels and the types of errors we are willing to risk (i.e., Type I or II) should be framed in a larger decision-making model. The focal point of this article and work is on *formal* decision-making and decision-making processes, systems and logics, more so than on a particular statistical method. In summary, the  $\chi^2$  model outlined here is a simple, practical, quick, low burden and easy to understand and execute approach that greatly improves the analysis of antibiogram data and decision-making by practitioners. More work and research is needed to develop additional inferential statistical methods and models that can be applied to antibiogram data.

## REFERENCES

1. Clinical Laboratory and Standards Institute/NCCLS. Analysis and presentation of cumulative antimicrobial susceptibility test data. Approved guideline-second edition. CLSI document M39-A2, 2005. NCCLS, 940 West Valley Road, Suite 1400, Wayne Pennsylvania 19087-1898.
2. Doern, G.V., A.B. Brueggemann, R. Perla, J. Daly, D. Halkias, R.N Jones and M.A. Saubolle, 1997. Multicenter laboratory evaluation of the bioMerieux Vitek antimicrobial susceptibility testing system with 11 antimicrobial agents versus members of the family *Enterobacteriaceae* and *Pseudomonas aeruginosa*. J. Clin. Microbiol., 35: 2115-19.
3. Steward, C.D., J. M. Mohammed, J.M. Swenson, S.A. Stocker, P.P. Williams, R.P. Gaynes, J.E. McGowan, Jr. and F.C. Tenover, 2003. Antimicrobial susceptibility testing of carbapenems: Multicenter validity testing and accuracy levels of five antimicrobial test methods for detecting resistance in *Enterobacteriaceae* and *Pseudomonas aeruginosa* isolates. J. Clin. Microbiol., 41: 351-8.

4. Sader, H.S., T.R. Fritsche and R.N. Jones, 2006. Accuracy of three automated systems (MicroScan WalkAway, Vitek and Vitek 2) for susceptibility testing of *Pseudomonas aeruginosa* against five broad-spectrum beta-lactam agents. *J. Clin. Microbiol.*, 44: 1101-04.
5. Munro, B.H., 2001. *Statistical Methods for Health Care Research*. Lippincott, Williams & Wilkins, Philadelphia.
6. Anderson, D.I., 2003. Persistence of antibiotic resistant bacteria. *Curr. Opin. Microbiol.*, 6: 452-6.
7. Stevens, J., 1986. *Applied multivariate statistics for the social sciences*. Erlbaum, Hillsdale, N.J.
8. Carifio, J. and W. Kermis, 1991. Alleged sampling biases: Towards a theory of sampling from large complex populations and the net randomizing effects of many simultaneous biasing factors. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.
9. Pier, G.B. and R. Ramphal, 2005. *Pseudomonas aeruginosa*. *Principles and Practice of Infectious Diseases* (Eds. G.L Mandell, J.E. Bennett and R. Dolin), pp: 2587-615. Elsevier, Philadelphia.
10. Courvalin, P., 2005. Antimicrobial drug resistance: Prediction is very difficult, especially about the future. *Emerg. Infect. Dis.*, 11: 1503-6.
11. Ferguson, G., 1989. *Statistical Analysis in Psychology and Education*. McGraw Hill, New York.