Robustness of the Chen–Dougherty–Bittner procedure against non-normality and heterogeneity in the coefficient of variation

Douglas A. Powell
National Cancer Institute at Frederick
Data Management Services
Frederick, Maryland 21702-1201

Lucy M. Anderson
Robert Y. S. Cheng
National Cancer Institute at Frederick
Laboratory of Comparative Carcinogenesis
Frederick, Maryland 21702-1201

W. Gregory Alvord
National Cancer Institute at Frederick
Data Management Services
Frederick, Maryland 21702-1201

Abstract. Chen, Dougherty, and Bittner [Y. Chen, E. R. Dougherty, and M. L. Bittner, *J. Biomed. Opt.* **2**(4), 364–374 (1997)] provided the derivation of a probability density function (PDF) for a signal ratio from a DNA microarray. This PDF is potentially useful for testing whether a pair of signals from the same gene has a common mean. The derivation of the PDF assumes the normality of all signal distributions and a common coefficient of variation (CV) for all signals within a microarray. The testing procedure requires the calculation of a common confidence interval for a microarray, based on a maximum likelihood estimator of the "common" CV, and the determination of whether or not a ratio for a particular gene falls within this interval. This study used Monte Carlo techniques and demonstrated that the procedure is *robust* to violations of normality and also to constancy in the coefficients of variation. A closer examination of the dynamics of the procedure found that the robustness was the result of offsetting effects. The size of the confidence interval was increased as a result of higher estimates of the common CV, as the actual CV pattern became heterogeneous. This effect mitigated the inflation in the size of the ratio as a result of increasing CV heterogeneity. These findings suggest that the Chen–Dougherty–Bittner procedure may be used even if underlying assumptions do not hold. © *2002 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1501561]

Keywords: DNA microarray technology; robustness; normality; coefficient of variation.

Paper JBO 02003 received Jan. 9, 2002; revised manuscript received Apr. 8, 2002; accepted for publication May 8, 2002.

1 Introduction

DNA microarray technology has emerged as a useful technology in genetic research. Using this technology, it is possible to quantitate nucleic acids through fluorescent intensities of signals from distinct tissue samples. The ratios of these signals can be calculated and inferences drawn as to which differences in signal intensity are meaningful.

Chen, Dougherty, and Bittner[1] studied the problem of analyzing the mRNA from two distinct tissue samples. They derived an elegant formulation for the probability density function of a signal ratio under the condition that the means of the two component signals are equal. The density function was derived with the assumption that signal expression levels across the gene population of a microarray are independent random variables. Further assumptions are that the signal expression levels follow a normal distribution and that all signal expression levels share a common coefficient of variation.

Informal discussions with biologists have suggested that the assumption of a constant coefficient of variation for the entire gene set may not always be appropriate. For example, in a recent study we utilized the microarray approach of Genome Systems, Inc., to measure the expression of RNA from female mouse liver tissues as it related to the level of serum corticosterone. Our data consisted of five pairs of samples that compared the livers of mice with high and low serum corticosterone, thus permitting the calculation of a coefficient of variation (CV) for each of the 8772 genes in the study. We found that the ranges for the CVs were substantial. For 28% of the genes, ratios from the five pairs of samples produced CVs in the range of 0.05–0.9. About 53% of the gene ratios were in the range from 1.0 to 4.9 and 19% were in the range from 5.0 to 9.9. Our experience is that wide ranges such as these are not unusual in studies involving large numbers of genes in complex biological systems. In this particular study, the 20 genes with the lowest expression intensities had an average CV of 6.5 ($\pm$2.6), a phenomenon commonly observed with low expression intensities. The 20 genes with the highest expression intensities, on the other hand, had a lower average CV of 1.0 ($\pm$1.9). Of the many genes in a large study, some may be tightly regulated while others may be inducible or repressible by factors that may not be related to experimental parameters.

Our concerns with the normality and constant coefficient of variation assumptions led us to conduct a *robustness* study

Address all correspondence to W. Gregory Alvord, Data Management Services, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201. Tel: 301-846-5101; Fax: 301-846-6196; E-mail: gwa@css.ncifcrf.gov

in connection with the Chen–Dougherty–Bittner (CDB) procedure. The question of robustness is an important issue with respect to any applied statistical technique. A *robust* procedure is one that is affected only slightly by appreciable departures from the assumptions involved. In regard to the CDB procedure the question of robustness reduces to the following: suppose the expression levels are *not* normally distributed and the coefficients of variation across the entire gene set are *not* constant. Can we still use the CDB procedure for constructing approximately accurate confidence intervals? How likely are we to draw erroneous conclusions about observed differences in expression ratios?

If the CDB procedure is robust, it may be used regardless of whether or not theoretical assumptions of normality and constancy in the coefficient of variation are satisfied. This paper reports results of Monte Carlo studies that assessed the impact of varying sample sizes, distribution shapes, and degrees of heterogeneity in coefficients of variation. We shall demonstrate that the Chen–Dougherty–Bittner procedure is unaffected by varying sample sizes and is, indeed, quite robust to even extreme deviations from normality and to substantial inequalities in the coefficients of variation.

2 Chen–Dougherty–Bittner Procedure

Chen et al.[1] developed an expression for the probability density function for a signal ratio. Assuming we have n genes represented in a microarray then two signal expressions for the kth gene are denoted as X_k and Y_k, respectively. For example, these two signals might come from treatment and control conditions, where X_k represents a signal for the kth gene obtained from a sample of mRNA taken from a treated cell and Y_k represents a signal for the kth gene taken from a cell in the control condition. Let the ratio of the two signals for the kth gene be denoted by T_k. Therefore

$$T_k = X_k / Y_k, \quad k = 1,...,n, \tag{1}$$

and the probability density function is derived for any T_k in the microarray under the null hypothesis that the means for X_k and Y_k are equal, i.e.,

$$H_0: \quad \mu_{X_k} = \mu_{Y_k} = \mu_k. \tag{2}$$

Assuming that all genes in the array have a common coefficient of variation, c, that the distributions of X_k and Y_k are $N(\mu_k, c^2 \mu_k^2)$, and that all signal measurements are independent random variables, the *approximate* probability density function for T_k is given by

$$f_{T_k}(t) = \{[(1+t)\sqrt{1+t^2}]/[c(1+t^2)^2\sqrt{2\pi}]\}$$
$$\times \exp[-(t-1)^2/2c^2(1+t^2)]. \tag{3}$$

The density function is approximate because no signal intensity can assume a negative value. Therefore, strictly speaking, signal measurements cannot be normally distributed. The portions of the normal curves that fall into the negative regions are taken to be negligible.

The probability density function (PDF) in Eq. (3) depends only on the common coefficient of variation c, and not on the parameters of the distributions of X_k and Y_k. Thus a test of the null hypothesis in Eq. (2) can be conducted at the α level of significance by integrating Eq. (3) and finding the $100*(\alpha/2)$ and the $100*(1-\alpha/2)$ percentile points of the distribution of T_k. Performing this integration, however, requires a value for c. This value is obtained by use of a maximum likelihood estimator, obtained by calculating the likelihood function across the entire gene array and then solving for $\hat{c}$ as the maximizing value. Thus c can be estimated from the data as follows:

$$\hat{c} = \sqrt{(1/n)\sum_{k=1}^{n} (t_k-1)^2/(1+t_k^2)}, \tag{4}$$

where n is the number of genes in the array and t_k is the observed signal ratio, calculated from the data.

With c estimated from the data using Eq. (4), the rejection region for testing, Eq. (2), can be established. Under the assumption of Eq. (2), approximately $\alpha*100\%$ of the t_k values would fall outside the region of rejection, i.e. $(1-\alpha)*100\%$ of the t_k values would be contained in the interval between the critical values for such a region of rejection. This assumes the distribution of the ratio behaves in accordance with Eq. (3).

3 General Method and Research Questions

The dependent variable of interest throughout this paper is the empirical containment rate (ECR) of the CDB procedure. The ECR is the proportion of times the CDB procedure correctly accepts the null hypothesis of equal signal means under specific sets of experimental conditions that are described below. The ECR is the complement of the empirical type I error rate (EER). The ECR and the EER must sum to 1.0. For example, suppose under a specific experimental setup, the proportion of times the procedure accepts the null hypothesis (the ECR) is 0.937. Then the proportion of times the procedure rejects the null hypothesis (the EER) is 0.063. Thus, in this example, ECR+EER=0.937+0.063=1.0. Since the ECR and the EER always sum exactly to 1.0, whatever factors explain the ECR also explain the EER.

The three independent variables that we consider are (1) sample size, the number of simulated pairs of signals per gene array, (2) distribution shape, the proximity of the simulated data to normality, and (3) pattern of coefficients of variation within a given gene array.

Three general questions are proposed and evaluated in this paper. (1) Is the CDB procedure robust with respect to (i) violation of the normality assumption and (ii) with respect to violation of the assumption of a common coefficient of variation across all genes in the array? (2) Can an adequate statistical model be developed to predict ECR from the independent variables in the study? (3) What accounts for the robustness or lack of robustness of this procedure?

The answer to the second question will determine to a great extent the answer to the first question. If a highly accurate predictive model can be developed for ECR, then the robustness or nonrobustness of the CDB procedure is inferable from such a model. In fact, the answer to the second question can even provide a more sensitive answer to the robustness question by informing researchers about what levels of deviation from basic assumptions do not cause substantial deviations from nominal (expected) rates of containment.

The answer to the third question involves an analysis of the dynamics of the CDB procedure to determine how it produces a decision to reject or not reject the null hypothesis. The procedure computes a ratio and then calculates a pair of critical values to be used as criteria for rejection. The latter depends on the proper estimation of the common coefficient of variation, c. Thus, non-normality and heterogeneous coefficients of variation can affect the procedure in two different ways: (1) by a possible *distortion* in the distribution of the ratio, or (2) by a possible effect on the *estimation* of c. The robustness, or lack thereof, of the CDB procedure will depend on the interplay of these two effects.

4 Study I: Method and Results

The first simulation study investigated the effects of the three independent variables on the empirical containment rate. The three independent variables are described in turn.

4.1 Sample Size

The sample size refers to the size of the microarray of paired signals. Four sample sizes were considered: 1000, 2000, 4000, and 8000. For a sample size of n, an array of n rows and two columns was generated, with the first column representing the X values and the second column representing the Y values.

4.2 Shape of Distribution

Four distribution shapes were simulated. These shapes were (1) extreme non-normality, skewness of 1.75 and excess kurtosis (i.e., kurtosis -3) of 3.75, (2) moderate non-normality, skewness of 1.00 and excess kurtosis of 1.00, (3) mild non-normality, skewness of 0.50 and excess kurtosis of 0.25, and (4) normality, skewness of 0.0 and excess kurtosis of 0.0.

The same distribution shape was induced for both the X and Y columns of each gene pair for a microarray with n rows. The distribution shape was induced by means of the Fleishman power method.[2] In this method, the desired distribution shape is induced by transforming a standard normal variate by means of a cubic transformation,

$$U = A + BZ + CZ^2 + DZ^3, \qquad (5)$$

where $Z \sim N(0,1)$, A, B, C, and D are constants corresponding to the particular skewness and kurtosis that is required, with the condition that $A = -C$, and U is a resulting random variate with mean equal to zero, variance equal to one, and with the required skewness and kurtosis.

4.3 Coefficient of Variation Pattern

Ten different patterns for the coefficients of variation of the X and Y signals were used. For each pattern, 10 different levels of coefficient of variation were employed for a given array of size n. One tenth of the rows in each array received a potentially different coefficient of variation. A pattern for an array is thus defined by the degree of variation in the coefficients of variation that was induced in the array. Selected patterns are presented in Table 1. The average coefficient of variation in each pattern is equal to 0.12 and the standard deviation within the pattern is calculated.

Once an array with n rows and two columns was generated with a given distribution shape throughout, the pattern of co-

Table 1 CV values for CV patterns: Study I.

Pattern	CV values[a]			SD of CVs
1	0.120	–	0.120	0.000
.				
.				
.				
4	0.090	–	0.150	0.021
.				
.				
.				
7	0.060	–	0.180	0.042
.				
.				
.				
10	0.030	–	0.210	0.063

[a] Mean CV for all patterns = 0.120.

efficients of variation was induced by multiplying these entries by the desired coefficient of variation and then adding one to both the X and Y entries in each row. Because the Fleishman transformation produces random variates with a mean of zero and a standard deviation of unity, the multiplication by the desired coefficient of variation and then the addition of one produces a random variate with mean equal to one, standard deviation equal to the desired coefficient of variation, and skewness and kurtosis values as desired.

4.4 Design of Simulation

The ratios (T's) were calculated by dividing the X entries by the Y entries for each row in the array. The estimated "common" coefficient of variation was obtained by means of Eq. (4). If a row contained a negative value for X and/or a negative value for Y, it was not used in the simulation.

Once the estimate of the common coefficient of variation ($\hat{c}$) was obtained, a 95% confidence interval was generated for the entire array. Each T value was checked to see if it was contained in this interval or not. The total proportion of valid rows (those containing both positive X and Y values) that had a T value contained within the 95% confidence interval was computed. This proportion represented the ECR for that array.

The design of the simulation thus constituted a completely crossed design with 4 sample sizes, 4 distribution shapes, and 10 coefficient of variation patterns. The resulting $4 \times 4 \times 10$ design produced 160 cells. For each cell in the design, 100 replicates (i.e., 100 arrays) were simulated. The dependent measure, the ECR, was computed for each replication within each cell. Thus a total of $160 \times 160 = 16\,000$ cases was generated in the simulation. All simulations were accomplished via programs written in the GAUSS programming language[3] and independently confirmed by programs written in S-PLUS.[4]

Table 2 Three-way fixed effects ANOVA model for Study I. Dependent variable: ECR. Weighted least squares.

Source	Degree of freedom	Sum of squares	Mean square	F statistic	p value of F
Sample	3	3.151	1.050	2.57	0.053
Shape	3	9900.964	3300.321	8077.68	<0.0001
CV Pattern	9	19 948.005	2216.445	5424.85	<0.0001
Sample Shape	9	6.333	0.704	1.72	0.078
Sample CV Pattern	27	11.069	0.410	1.00	0.459
Shape CV Pattern	27	1402.131	51.931	127.10	<0.0001
Sample Shape CV Pattern	81	29.225	0.361	0.88	0.764
Error	15 840	6471.792	0.409		

4.5 Analysis of Results

The data from the simulation were analyzed by means of a three-way fixed effects analysis of variance (ANOVA). The analysis[5] was done via weighted least squares with ECR being the dependent variable and the three independent variables previously described as factors. Each ECR was weighted by the reciprocal of its respective sampling variance to compensate for the heterogeneity of variances that is inherent in the analysis of proportions.

For this analysis the ECR was multiplied by $(1-\text{ECR})$ and then divided by the number of valid rows for the gene array to establish the sampling variance for the estimated proportion. The results for this analysis are given in Table 2.

From this analysis a preliminary step towards a general model can be made. Sample size shows no statistically significant relationship with ECR, either in terms of main effects or interaction effects. On the other hand, the distribution shape and CV pattern both show significant main effects and a significant interaction effect. The results suggest a two-way model for ECR with both CV pattern and distribution shape playing explanatory roles.

5 Study II: Method and Results

To further investigate the relationship between non-normality, CV pattern, and ECR, a more extensive study was done. Given the lack of a significant relationship between sample size and ECR, the data for all cells was simulated with a sample size of 8000. The study also took into account the fact that an interaction effect was found between the CV pattern and the distribution shape in study I.

5.1 General Procedure

To investigate the fact that the relationship among CV pattern and ECR is influenced by distribution shape, a separate simulation was done for each of the four distribution shapes that were employed in study I. Thirty different patterns for the

coefficient of variation were used in this second study. Selected CV patterns are given in Table 3. We note that the average CV in each pattern is 0.15. Each of the 30 CV patterns was simulated 1000 times for each of the four distribution shapes. This led to 30 000 ECRs for each distribution shape. The simulations were done by means of programs written in the GAUSS computer language[3] and results were independently confirmed by a set of programs written in S-PLUS.[4]

In addition to the ECR, the quantity ECR15 was computed. The ECR15 statistic is the rate of containment in a 95% confidence interval for the signal ratio, assuming the coefficient of variation is fixed at 0.15. Thus the ECR is computed by the actual estimate of the assumed to be the common coefficient of variation using Eq. (4). The ECR15 is computed by fixing the CV at the average CV level for each pattern, i.e., 0.15. More precisely, the ECR15 was computed by specifically using the lower and upper 95% confidence limits for $c=0.15$.

5.2 Assessment of Robustness

The total impact of the variation in distribution shape and CV pattern is defined as the difference between 0.95 and the ECR. Thus, the *total effect*, TE, of variation in both the CV pattern and distribution shape is given by

$$\text{TE}=0.95-\text{ECR}. \qquad (6)$$

A positive value of the TE reflects a rate of containment less than 0.95, i.e., a rate of rejection in excess of 0.05. A negative value of the TE reflects a rate of containment greater than 0.95, i.e., a rate of rejection less than 0.05.

The ECR is a composite of two separate effects. The CDB procedure estimates the assumed to be the common coefficient of variation, c, and produces the estimate $\hat{c}$. It then uses this value to calculate a confidence interval that either contains the calculated value of T or does not. Thus, the ECR is in part affected by the size of this confidence interval due to the value of $\hat{c}$. It is also affected by any distortion in the distri-

Table 3 CV values for CV patterns: Study II.

Pattern	CV values[a]			SD of CVs
1	0.150	–	0.150	0.000
.				
.				
.				
8	0.115	–	0.185	0.024
.				
.				
.				
15	0.080	–	0.220	0.049
.				
.				
.				
23	0.040	–	0.260	0.077
.				
.				
.				
30	0.005	–	0.295	0.101

[a] Mean CV for all patterns=0.150.

bution of T itself. The TE might thus be a good "bottom-line" statistic in assessing robustness, but it does not tell us why the ECRs deviate from the nominal value of 0.95.

Two components of TE are now defined. The *distortion effect* (DE) reflects the amount of deviation of the empirical containment rate from the nominal value of 0.95 when the size of the confidence interval is held constant. In this study, the size of the confidence interval was held constant by using an interval based on a CV equal to 0.15. The DE is thus given by

$$DE = 0.95 - ECR15. \qquad (7)$$

A positive value of DE reflects a rate of containment less than 0.95 and a negative value reflects a rate of containment greater than 0.95, with the rates of containment being computed by holding the size of the confidence interval constant. Variation in the size of DE reflects the pure effect of the distortion of the distribution of T because of different CV patterns and different distribution shapes.

A second component of TE reflects the difference in the rate of containment with the fixed confidence interval size and the rate of containment based on a variable confidence interval size. The variation in confidence interval size is a function of the variation in the estimated common coefficient of variation. This effect is termed the *estimation effect* (EE) and is defined as follows:

$$EE = ECR15 - ECR. \qquad (8)$$

A positive value of EE reflects a higher rate of containment with the fixed confidence interval as opposed to the confidence interval deriving from the estimation of the CV. A negative value of EE reflects a lower rate of containment with the fixed confidence interval size as opposed to the confidence interval deriving from the estimation of the CV. It is true that the larger the estimated CV the larger the size of the corresponding confidence interval.[1] Thus, larger estimated CVs are associated with higher ECRs and smaller EEs.

Combining Eqs. (6)–(8) permits the statement of the following relationship:

$$TE = DE + EE. \qquad (9)$$

Robustness can thus be conceived of as the pattern of behavior of TE. If TE is always zero, we would have "perfect" robustness. If TE has an average value far removed from zero under conditions where either distribution shape deviates from normality or the CV pattern deviates from equality of CV, then it can be concluded the CDB procedure is not robust.

Furthermore, by using the relationship in Eq. (9) the question of why the procedure is robust or not robust may be addressed. By examining the patterns of DE and EE in addition to that of TE, one may determine whether a finding of robustness is due to comparably small absolute average DEs and EEs across different experimental conditions, or is due to large absolute DEs and EEs that are moving in opposite directions. Likewise, a finding of nonrobustness might be due to near zero EEs and huge DEs or vice versa, or moderately large departures from zero for both DEs and EEs.

5.3   Results

To demonstrate that the data were generated according to specifications, the average values of the mean, variance, skewness, and excess kurtosis are reported for both the X and Y variables, for each distribution shape, at the CV pattern that involves complete homogeneity of CV. These results are presented in Table 4. The four moments of the distribution are very close on average to the intended values, although the Fleishman procedure slightly underestimates the excess kurtosis.

The average estimated coefficients of variation by the CV pattern are displayed in Figure 1 for each distribution shape. The average estimated CV increases as the heterogeneity of the CV pattern increases. This relationship holds for all distribution shapes, although in the three non-normality conditions, the common CV of 0.15 is underestimated under the experimental condition of complete homogeneity of the CVs. The degree of underestimation is in direct proportion to the non-normality of the distribution.

The average, minimum, and maximum values of ECR for each CV pattern by distribution shape are given in Table 5. A quick summary of the findings shows that of the 120 000 replications in the entire study, the lowest ECR observed was 0.916 (minimum for CV pattern 30 with Extreme NN) and the highest ECR observed was 0.959 (maximum for CV pattern 1 with Mild NN). For the condition combining the most serious deviation from normality (Extreme NN) and the most extreme heterogeneity of coefficients of variation (CV pattern 30), the

Table 4 Average mean, variance, skew, and excess kurtosis by distributional shape, CV pattern reflecting all CVs equal to 0.15. (Averages are based on 1000 replications.)

Statistic	Normal	Mild NN	Moderate NN	Extreme NN
Avg mean X	0.999 97	1.000 07	1.000 09	1.000 03
Avg variance X	0.022 48	0.024 90	0.022 51	0.022 49
Avg skew X	0.000 60	0.496 73	0.992 99	1.733 32
Avg xkur X[a]	−0.007 67	0.238 94	0.968 38	3.633 37
Avg mean Y	1.000 01	1.000 09	0.999 95	0.999 93
Avg variance Y	0.022 50	0.022 51	0.022 48	0.022 48
Avg skew Y	−0.000 18	0.495 17	0.992 66	1.737 66
Avg xkur Y[a]	−0.007 14	0.232 05	0.965 50	3.639 52

[a] xkur=excess kurtosis.

average ECR was 0.922. The ECRs are represented graphically in Figure 2. Also represented in Figure 2 are the ECR15s. Note that for the CV pattern reflecting complete homogeneity [standard deviation (SD) of CV=0] and with a normal distribution of the signals, both the ECR and the ECR15 are precisely equal to 0.95. The two sets of containment rates exhibit the same decreasing pattern for all distribution shapes.

TEs, DEs, and EEs were computed and are represented graphically in Figure 3 as a function of the heterogeneity of the CV pattern. The EE shows a decreasing pattern for all distribution shapes. This reflects the fact that the estimated coefficient of variation increases with the heterogeneity in the CV pattern. Likewise, the displacement in the graphs of EE by the distribution corresponds to the displacement in the graphs of the estimated coefficients of variation by distribution shape. The pattern in the DE shows an opposing pattern to that in the EE. DE increases for every distribution shape as the heterogeneity of the CV pattern increases. Furthermore, under the homogeneity of the CV condition, there was no observed distortion effect for the normal distribution shape but there were negative distortion effects for each of the non-
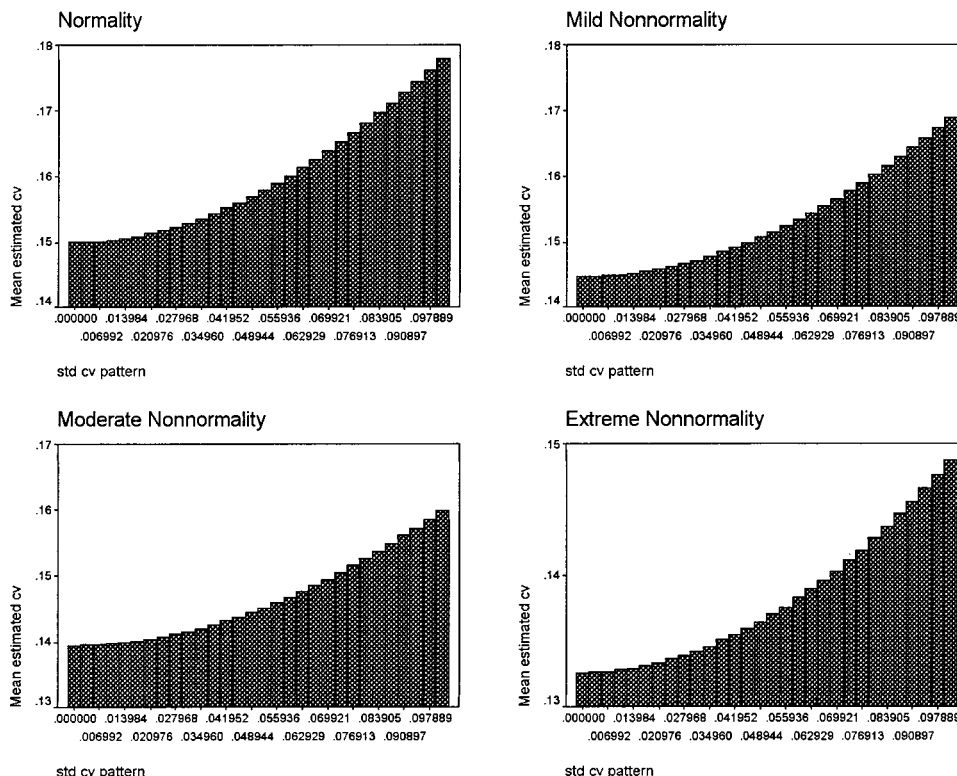

Fig. 1 Mean estimated CV by distribution shape.

Table 5 Average, minimum, and maximum ECR by CV pattern by distribution shape (based on 1000 replications).

CV Pattern	Normal			Mild NN			Moderate NN			Extreme NN		
	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
1	0.950	0.945	0.955	0.953	0.948	0.959	0.951	0.946	0.956	0.937	0.932	0.942
2	0.950	0.946	0.955	0.953	0.946	0.958	0.951	0.945	0.956	0.937	0.932	0.943
3	0.950	0.944	0.955	0.953	0.947	0.959	0.951	0.945	0.956	0.937	0.932	0.944
4	0.950	0.945	0.955	0.953	0.947	0.958	0.950	0.945	0.955	0.937	0.932	0.943
5	0.949	0.944	0.954	0.952	0.947	0.957	0.950	0.944	0.954	0.937	0.931	0.943
6	0.949	0.943	0.954	0.951	0.945	0.957	0.949	0.944	0.954	0.937	0.931	0.943
7	0.948	0.943	0.953	0.951	0.946	0.956	0.949	0.944	0.954	0.936	0.931	0.941
8	0.948	0.941	0.953	0.950	0.944	0.956	0.948	0.942	0.953	0.936	0.931	0.942
9	0.947	0.940	0.952	0.949	0.944	0.954	0.947	0.940	0.952	0.936	0.930	0.941
10	0.946	0.940	0.952	0.948	0.942	0.953	0.946	0.941	0.951	0.936	0.930	0.942
11	0.945	0.941	0.951	0.946	0.941	0.951	0.945	0.939	0.952	0.935	0.930	0.940
12	0.944	0.938	0.950	0.945	0.939	0.950	0.944	0.937	0.949	0.935	0.930	0.941
13	0.943	0.937	0.949	0.944	0.938	0.949	0.943	0.937	0.948	0.934	0.928	0.940
14	0.942	0.937	0.947	0.943	0.936	0.948	0.941	0.935	0.947	0.933	0.928	0.938
15	0.941	0.936	0.946	0.941	0.936	0.946	0.940	0.935	0.945	0.933	0.927	0.939
16	0.940	0.934	0.945	0.940	0.935	0.944	0.939	0.933	0.944	0.932	0.927	0.938
17	0.939	0.932	0.944	0.938	0.932	0.943	0.939	0.932	0.943	0.931	0.926	0.937
18	0.938	0.933	0.944	0.937	0.931	0.941	0.936	0.931	0.943	0.931	0.925	0.937
19	0.937	0.931	0.942	0.936	0.931	0.943	0.935	0.930	0.941	0.930	0.923	0.934
20	0.936	0.930	0.941	0.935	0.930	0.940	0.934	0.929	0.939	0.929	0.924	0.934
21	0.935	0.930	0.940	0.933	0.928	0.938	0.933	0.927	0.938	0.928	0.923	0.933
22	0.934	0.929	0.940	0.932	0.927	0.938	0.931	0.926	0.937	0.927	0.922	0.933
23	0.933	0.928	0.939	0.931	0.924	0.937	0.930	0.924	0.936	0.927	0.921	0.932
24	0.932	0.928	0.938	0.930	0.924	0.936	0.929	0.924	0.934	0.926	0.919	0.931
25	0.931	0.926	0.937	0.929	0.924	0.934	0.928	0.922	0.934	0.925	0.920	0.930
26	0.931	0.926	0.936	0.928	0.923	0.934	0.927	0.923	0.932	0.924	0.919	0.930
27	0.930	0.925	0.936	0.928	0.923	0.932	0.926	0.922	0.931	0.924	0.916	0.929
28	0.929	0.924	0.936	0.927	0.921	0.932	0.926	0.921	0.931	0.922	0.917	0.929
29	0.929	0.923	0.934	0.926	0.920	0.931	0.925	0.919	0.931	0.922	0.917	0.927
30	0.928	0.923	0.934	0.925	0.920	0.930	0.924	0.919	0.930	0.922	0.916	0.927
Overall	0.940	0.923	0.955	0.940	0.920	0.959	0.939	0.919	0.956	0.931	0.916	0.944

normal shapes. As a general rule, the DE is more positive as the distribution shape gets closer to normality.

The moments of the signal ratio distributions are displayed in Table 6. The mean, variance, skewness, and kurtosis all increase as a function of CV heterogeneity. Both Pearson and Spearman correlations between these moments and distortion effects are also presented in Table 6. Distortion effects are strongly positively correlated with the four distribution indices. In the case of the ratios under the normality condition, the clear effects of extreme outliers are observed both in the case of the variance of the ratio and in the case of the Pearson correlation with DE. The nonparametric Spearman correlation
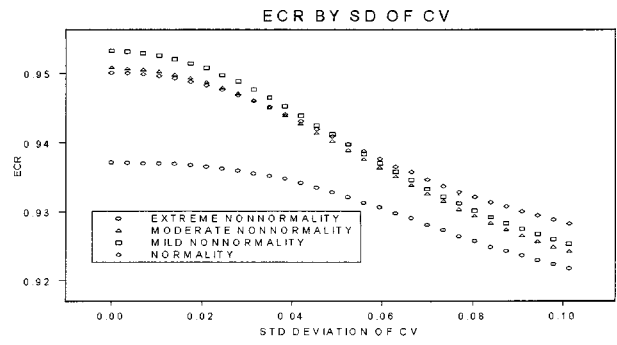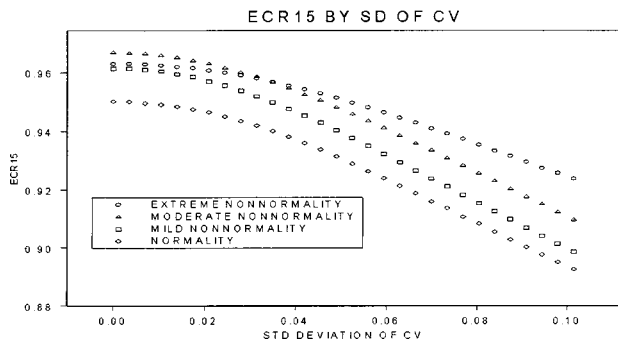
Fig. 2 Average ECR and ECR15 by SD of CV.

was therefore included to provide an indicator of the relationship of the ratio variance and the distortion effect that is not sensitive to extreme values.

The increases in the moments of the ratio distribution as a function of CV heterogeneity are moderated by the distribution shape of the constituent signals. If we take the moments of the ratio distribution for the normal condition and CV homogeneity as normative, it is clear that all of the non-normal shapes show a smaller dispersion in the signal ratio at low levels of CV heterogeneity than this normative level. This accounts for the negative distortion effects in the non-normal conditions at low levels of heterogeneity of CV. The signal ratio dispersion increases as the CV heterogeneity increases, and the differences in ratio dispersion among distribution shapes of the constituent signals are consistent with the differences in the graphs of DE by distribution shape of the signals.

5.4 Modeling of ECR and ECR15

The graphs of ECR and ECR15 (see Figure 2) suggest a nonlinear regression model for containment rates. In particular the following four-parameter logistic model[6] was employed:

$$CR_i = (A - D)/(1 + S_i/C)^B + D + e_i, \qquad (10)$$

where CR_i is the containment rate for the ith CV pattern (either ECR or ECR15), S_i represents the standard deviation of the ith CV pattern, parameter A represents the containment rate when the standard deviation of the CV pattern equals zero, D is the containment rate assuming infinite variation in the pattern of CVs, C is the standard deviation of the CV pattern that produces a containment rate exactly halfway between A and D, and B represents a "slope factor" that deter-
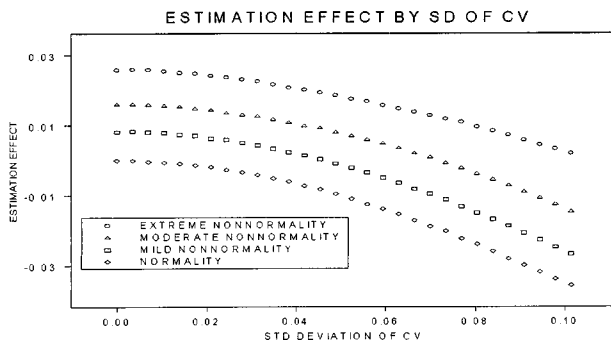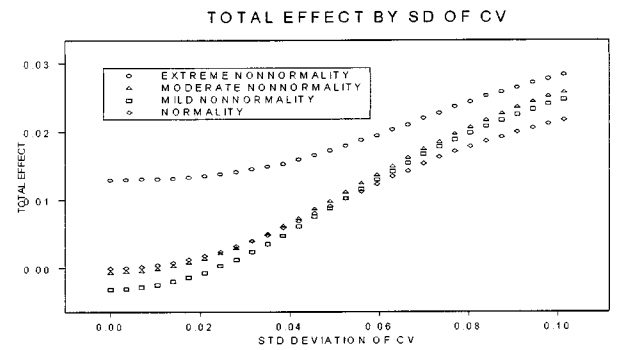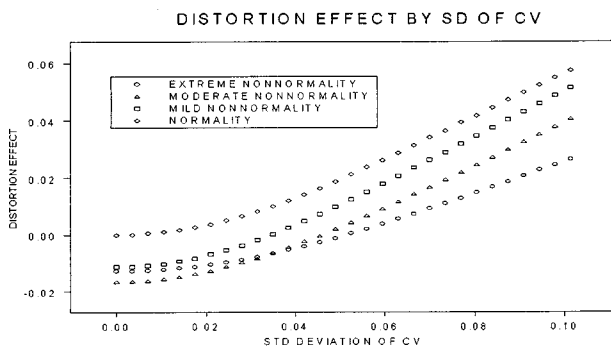

Fig. 3 Average distortion effect, estimation effect, and total effect by SD of CV.

Table 6 Average mean, average variance, average skew, and average kurtosis of ratio by selected-CV patterns. (All statistics are based on 1000 replications, and correlations are based on a sample size of 30 000.)

Pattern	Normality				Mild non-normality			
	Avg. mean	Avg. variance	Avg. skew	Avg. kurtosis	Avg. mean	Avg. variance	Avg. skew	Avg. kurtosis
1	1.0242	0.052	0.791	4.723	1.0223	0.047	0.612	3.547
4	1.0244	0.052	0.828	4.965	1.0225	0.047	0.629	3.643
7	1.0246	0.054	0.943	5.937	1.0228	0.048	0.689	3.978
10	1.0257	0.056	1.230	13.628	1.0233	0.050	0.778	4.487
13	1.0269	0.060	1.511	14.867	1.0241	0.052	0.906	5.235
16	1.0282	0.069	2.454	60.083	1.0252	0.055	1.054	6.115
19	1.0303	0.084	3.687	111.489	1.0263	0.059	1.225	7.199
22	1.0365	97.865	6.918	313.816	1.0282	0.063	1.411	8.462
25	1.0368	1.187	10.438	535.867	1.0300	0.068	1.612	9.898
28	1.0434	22.511	14.630	821.201	1.0318	0.075	1.820	11.562
30	1.0444	5.148	16.541	927.335	1.0338	0.080	1.993	13.094
r[a]	0.161[c]	0.013[d]	0.410[c]	0.295[c]	0.788[c]	0.977[c]	0.966[c]	0.919[c]
rho[b]	0.841[c]	0.969[c]	0.921[c]	0.899[c]	0.770[c]	0.989[c]	0.973[c]	0.968[c]
Pattern	Moderate non-normality				Extreme non-normality			
	Avg. mean	Avg. variance	Avg. skew	Avg. kurtosis	Avg. mean	Avg. variance	Avg. skew	Avg. kurtosis
1	1.0207	0.043	0.609	3.596	1.0188	0.040	0.816	4.850
4	1.0208	0.044	0.626	3.683	1.0190	0.040	0.834	4.961
7	1.0209	0.044	0.672	3.938	1.0191	0.041	0.883	5.281
10	1.0213	0.046	0.746	4.348	1.0194	0.042	0.958	5.762
13	1.0221	0.047	0.843	4.892	1.0200	0.043	1.064	6.450
16	1.0228	0.049	0.957	5.529	1.0206	0.045	1.183	7.207
19	1.0238	0.052	1.077	6.195	1.0213	0.047	1.311	8.049
22	1.0252	0.055	1.217	6.993	1.0223	0.049	1.450	8.930
25	1.0264	0.059	1.343	7.715	1.0233	0.052	1.582	9.794
28	1.0282	0.063	1.476	8.502	1.0245	0.056	1.710	10.647
30	1.0291	0.067	1.563	9.030	1.0253	0.058	1.799	11.259
r[a]	0.723[c]	0.982[c]	0.971[c]	0.958[c]	0.651[c]	0.979[c]	0.948[c]	0.925[c]
rho[b]	0.705[c]	0.987[c]	0.971[c]	0.968[c]	0.631[c]	0.980[c]	0.948[c]	0.939[c]

[a] Pearson's correlation between the distortion effect and the variable.

[b] Spearman's correlation between the distortion effect and the variable.

[c] $p < 0.01$.

[d] $p < 0.05$.

mines the slope of the curve. The larger the value of B, the steeper the curve. The e_i term reflects random errors, assumed to be normally distributed (see Table 7).

The models for ECR15 reflect the phenomenon of increasing distortion effects as a function of CV heterogeneity. Esti-

mates of parameter A accurately reflect ECR15 for each distribution shape under the CV homogeneity condition. Likewise, estimates for parameter D indicate the impact of CV heterogeneity on the containment rates via the mechanism of signal ratio distortion. The fact that the parameter D esti-

Table 7 Four-parameter logistic model results for both ECR and ECR15. R^2_{logist} is the proportion of variance in containment rates explained by logistic model. R^2_{ANOVA} is the proportion of variance in containment rates explained by a one way ANOVA with the CV pattern as the grouping variable, i.e., the theoretically best fitting model for the containment rate.

	ECR							
	Normal		Mild NN		Moderate NN		Extreme NN	
Parameter	Est.	SE	Est.	SE	Est.	SE	Est.	SE
A	0.9500	0.00002	0.9531	0.00002	0.9507	0.00002	0.9370	0.00002
B	2.3668	0.01843	2.2692	0.01421	2.2807	0.01553	2.7921	0.03361
C	0.0707	0.00052	0.0717	0.00046	0.0783	0.00061	0.0870	0.00110
D	0.9190	0.00026	0.9126	0.00028	0.9093	0.00036	0.9117	0.00045
R^2_{logist}	0.95116		0.96835		0.96469		0.89848	
R^2_{ANOVA}	0.95123		0.96840		0.96472		0.89865	
	ECR15							
	Normal		Mild NN		Moderate NN		Extreme NN	
Parameter	Est.	SE	Est.	SE	Est.	SE	Est.	SE
A	0.9502	0.00004	0.9616	0.00004	0.9669	0.00003	0.9631	0.00003
B	2.0349	0.01247	1.9642	0.01015	1.9338	0.01075	2.0337	0.01648
C	0.1263	0.00193	0.1293	0.00177	0.1469	0.00259	0.1525	0.00399
D	0.8028	0.00270	0.7975	0.00263	0.7930	0.00370	0.8334	0.00439
R^2_{logist}	0.97683		0.98312		0.98121		0.96262	
R^2_{ANOVA}	0.97686		0.98314		0.98124		0.96266	

mates for the ECR15 model are substantially lower than those for the ECR model is a testament to the "compensatory" effect on ECR resulting from the estimation of the coefficient of variation.

6 Discussion

This study was designed to answer three questions. First we asked, "Is the CDB procedure robust with respect to (1) deviations from normality in the shapes of the signal distributions and to (2) deviations from homogeneity of the coefficients of variation of the signal distributions?" The answers to these questions are "Yes." The worst case of a deviation from 0.950 was 0.916 in 120 000 replications. Each of the distribution shapes had an average ECR between 0.922 and 0.928 under even the worst pattern of CV heterogeneity. For most cases of CV heterogeneity the resulting ECRs were even closer to the nominal containment rate of 0.950.

In practice this means that even if expression levels in microarrays are *not* normally distributed and even if the coefficients of variation across the gene set are *not* constant, the Chen–Dougherty–Bittner procedure may still be used for constructing highly accurate confidence intervals. Even under extreme deviations from the assumptions of Chen, Dougherty,

and Bittner, an error in rejecting the null hypothesis that we would expect to occur only 5% of the time, would, in reality, occur no more often than 8% of the time.

Another question posed was, "Can an adequate explanatory model be developed to predict the ECR from the independent variables in the study?" The answer is, again, "Yes." Sample size played no role in explaining empirical containment rates. The ECR decreased as the CV heterogeneity increased. Distribution shape moderated this relationship. In general, the ECR was closest to the nominal level of 0.95 for the normality condition and farthest away from 0.95 for the extreme non-normality condition. The results for the two intermediate non-normality conditions showed similarity to each other. Each had ECR values that were too high under CV homogeneity. However, ECR values dipped below those for the normality condition as the CV heterogeneity increased.

A four-parameter logistic model was used to describe the data. It showed an outstanding fit to the data for all cases of non-normality. The model permitted an estimate of what the ECR would be if the CV heterogeneity were to become infinitely large. The results were consistent with the previously expressed finding that the CDB procedure is robust against violations of its relevant assumptions.

A final question involved an assessment of why the CDB procedure is robust. The decision as to whether or not to reject the null hypothesis of equality of means has two parts. A ratio is computed and a confidence interval is obtained. If the ratio is in the interval, the null hypothesis is not rejected. Otherwise it is rejected. Thus the rate of containment of the procedure is a function of the distribution of the signal ratio and the size of the confidence interval. The size of the confidence interval is strictly a function of the estimate of the common coefficient of variation.

The TE was shown to be the sum of the EE and the DE. We found that, for each distribution shape, the two effects offset each other. For normal data, large positive distortion effects are greatly offset by large negative estimation effects, thus stabilizing the total effects and yielding ECRs that are very close to the nominal level. Other distribution shapes show a similar pattern of rising distortion effects accompanied by falling estimation effects. The non-normality conditions have negative distortion effects offset by positive estimation effects at low CV heterogeneity levels. However, the trends in the graphs are the same for each distribution shape. For the extreme non-normality condition, the actual level of the EE remains positive over all CV patterns. The two effects of CV heterogeneity have opposite directions. They serve to mitigate each other. Increased dispersion in the ratio distribution is checked by increases in the estimated common coefficient of variation, leading to concomitantly larger confidence intervals. The net effect is relative stability in the ECR.

In summary, we have demonstrated that the Chen–Dougherty–Bittner procedure is robust to deviations of assumptions regarding normality and constancy in the coefficients of variation in the signal distributions. We have developed a statistical model that explains how the empirical containment rate would be affected if heterogeneity in the coefficient of variation were to become infinitely large. Finally, we have shown that distortion effects and estimation effects offset each other in contributing to total effects on empirical containment rates, further explaining the reasons for the robustness of the procedure.

References

1. Y. Chen, E. R. Dougherty, and M. L. Bittner, ''Ratio-based decisions and the quantitative analysis of cDNA microarray images,'' *J. Biomed. Opt.* **2**(4), 364–374 (1997).
2. A. Fleishman, ''A method for simulating nonnormal distributions,'' *Psychometrika* **43**, 4 (1978).
3. Gauss System Command Reference Version 3.0. Aptech Systems, Inc., Maple Valley, WA (1992).
4. S-PLUS 2000 Programmer's Guide, MathSoft, Inc., Seattle, WA (1999).
5. SAS User's Guide: Statistics, Ver. 5 edition, SAS Institute, Inc., Cary, NC (1985).
6. A. DeLean, P. J. Munson, and D. Rodbard, ''Simultaneous analysis of families of sigmoidal dose-response curves: Application to bioassay, and physiological dose-response curves,'' *Am. J. Physiol.* **235**, E97 (1978).