

The use of multivariate regression analysis in contrast-detail studies of CT scanners

Manuel Trajtenberg^{a)}

Harvard University and The National Bureau of Economic Research, Cambridge, Massachusetts 02138

(Received 29 June 1983; accepted for publication 23 November 1983)

Previous studies of the imaging performance of computed tomography (CT) scanners, and other imaging modalities, have failed to apply appropriate statistical methods to data analysis, thus impairing the accuracy and significance of results. Given that imaging performance involves a number of interrelated variables and an element of randomness, its empirical assessment requires multivariate regression analysis. This method is used here to analyze anew a set of contrast-detail data from a previous study on CT scanners. The main issues considered are the specification of the proper functional form linking perceptibility, dose and contrast, the estimation of the contrast and dose coefficients, and of scanner-specific constants to be used in computing indices of imaging quality. One of the main empirical findings is that the dose coefficient of the CT scanners studied is significantly less than that predicted by the theoretical model: 1/5 instead of 1/3. This result suggests that actual dose used in routine clinical studies could be reduced substantially without impairing much the quality of the images. On the other hand, the coefficient of contrast does correspond to its predicted value, i.e., 2/3. The methodology used here is not limited to the contrast-detail framework, but is applicable to, and indeed essential in, empirical studies of the performance of any imaging modality.

I. INTRODUCTION

The main purpose of this article is to introduce and urge the application of the statistical methodology associated with multivariate regression analysis,^{b)} to the empirical assessment of image quality of computed tomography (CT) scanners and, by extension, of other imaging modalities as well.

Following the emergence, constant improvement and rapid proliferation of a vast array of new imaging technologies in the past decade, it has been a priority concern of radiologists and medical physicists to devise acceptable means and standards for assessing the performance of these systems. A great deal of work has been done for that purpose (particularly in CT), but there seems to be a serious imbalance in the components of such research: whereas it does have well established theoretical foundations, and is fairly rigorous—and imaginative—in the acquisition of data, it exhibits severe deficiencies in the choice and use of statistical methods for data analysis. Not surprisingly, this has led in some cases to erroneous conclusions, and has definitely impaired the comparability of results obtained in different studies. The main problems are

(i) The exclusive use of *bivariate* rather than *multivariate* statistical methods in the analysis of phenomena that encompass, in theory and practice, a number (> 2) of interrelated variables. For example, the estimation of partial (pairwise) correlations between, say, detectability and contrast, or detectability and dose, instead of estimating the structural relationship between all of these and related variables that jointly determine the imaging behavior of a system. The most severe consequence of this type of "specification error" is that the estimated coefficients (e.g., the dose coefficient of CT scanners) are most likely to be *biased*, and the same is true for the corresponding measures of statistical signifi-

cance (e.g., standard errors).

(ii) The failure to "pool" (that is, to estimate jointly) the data on the various systems being assessed, thus precluding the proper estimation of "figures of merit" for summary comparisons among systems.^{c)}

(iii) More fundamentally, some of the applied research done in this area has either ignored altogether, or failed to recognize that theoretical models (i.e., mathematical functions) describing the properties of imaging systems are idealized constructs to be used only as starting hypotheses: it is the task of the statistical analysis of empirical data to reveal the actual magnitudes of the coefficients, and the appropriate mathematical forms.

I shall attempt to show the significance of these and related methodological issues, by analyzing anew data on "contrast-detail-dose" of CT scanners, with the aid of multivariate regressions. These data were originally gathered and analyzed by Cohen¹ and Cohen and DiBianca.² Subsequently, Dr. Cohen kindly agreed to provide me with their data in order to perform this study. As suggested above, the approach put forward here can, and indeed should, be applied not only to CT scanners, but to any imaging modality. Concrete examples of studies that call for such application include: Wagner *et al.* on radiographic magnification,³ Gould *et al.* on intensified fluoroscopic images,⁴ and Smith and Lopez on Ultrasound.⁵

II. THE THEORETICAL MODEL: INTERPRETATION AND DOMAIN

The mathematical relationship underlying studies of image quality of CT scanners can be stated as follows (for the standard derivation see, for example, Ref. 6):

$$Dw^3(\Delta\mu)^2 = F, \quad (1)$$

where D represents dose, F is a system-specific constant, and w and $\Delta\mu$ have been interpreted in a variety of ways: Hounsfield⁷ and Bassano *et al.*⁸ for example, take $\Delta\mu$ (σ in their notation) to mean "noise" (or "inaccuracy", or "grain") and w actual pixel size, i.e., the object diameter divided by the number of rows in the matrix. In that context noise is usually measured by the percentage standard deviation of the CT numbers in a given area of a water bath or a similar statistic. Southon⁹ also interprets $\Delta\mu$ to be noise, but takes w to be spatial resolution (Δr in his notation), calculated from scans of air holes and making use of modulation transfer functions (MTFs) in a rather nonconventional way. A significantly different interpretation is that adapted in "contrast-detail" studies such as in Refs. 1 and 2: they take $\Delta\mu$ to be contrast, predetermined in the construction of the phantom (its nominal value at least) and w (d in their notation) to be "threshold detectability", i.e., the diameter of the smallest detectable dot in the scan, to be assessed by eye inspection. The two interpretations are, of course, equally legitimate, and represent different aspects of the same underlying phenomena, connected via the concept of signal-to-noise ratios (see for example Wagner *et al.*¹⁰). However, it remains to be seen whether the two are also empirically equivalent (except for a proportionality factor),^{d)} i.e., whether they will produce essentially the same assessment of imaging performance, or at least comparable/consistent results.

The second issue has to do with the *domain* of the model, that is, the range of values of the independent variables over which the model is expected to hold, and the related issue of its proper functional form. If taken literally, Eq. (1) implies that for a sufficiently large pixel size and/or dose, "noise" can be made infinitesimally low or, conversely, that spatial resolution can be increased as much as desired. But that is clearly unfeasible: in the context of the "contrast-detail" interpretation, for example, spatial resolution is ultimately bounded by the MTF, and contrast by the intrinsic or "floor" noise of the system. The question is thus whether the model should be regarded as essentially discontinuous, holding as in Eq. (1) but only within a reduced domain (to be determined) or as smooth and continuous, in which case both the functional form in Eq. (1) and its coefficients are at issue.

Finally, and regardless of the particular interpretation given to it, there is the question of the proper use of the model in empirical research. As stated above, a theoretical model purporting to describe a given phenomenon cannot be taken for granted, but has to be tested and verified empirically. Until that is done repeatedly and convincingly, it can only be regarded as a tentative guideline for applied research. This is particularly true in the case at hand: Equation (1) is in fact a description of an *ideal* CT system, based solely on the behavior of x rays. But the actual imaging performance of a system depends upon a host of factors that are not accounted for in the model, such as filter function, detector geometry (and fluctuations in their response), collimation efficiency, reconstruction algorithms, etc., etc. Thus, it is imperative to *estimate* the model with actual data, and use methods of statistical inference to assess both its coefficients [i.e., the exponents of Eq. (1)] and its functional form.

Most existing works have bypassed this critical stage. Some, like White *et al.*¹² have acknowledgedly ignored the model altogether, while others have taken it for granted. Southon,⁹ for example, calculated "figures of merit" for each scanner (and mode of operation) simply by plugging in the model the observed values of the variables, thus assuming the problem away. Cohen¹ did attempt to estimate it (at least partially), but there are problems regarding the adequacy of his statistical procedures.

III. STATISTICAL ESTIMATION

A. The data set

As said above, the data used here were furnished to me by Dr. G. Cohen and are essentially the same that he used in his study.¹ The only exceptions are that the EMI Mark I was not included here (its data were unavailable) and, whereas in Ref. 1 only 2 scans for the GE7800 were considered, I used here also the data on the GE7800 from Ref. 2, comprising six additional scans. Although I assume otherwise familiarity with Cohen's paper, I shall briefly describe what these data consist of (see Table I). Six different scanners were studied, under a variety of operational modes: the EMI CT 1005 (hereafter EMI), the Pfizer 200FS (Pfizer), the AS&E stationary detector system (AS&E), the Searle Photrax 4000 (Searle), the General Electric CT/T 7800 (GE78), and the General Electric 8800 (GE88). The scanning parameters common to all scanners include kVp, slice thickness, scan time, mA, and peak and summation dose. However, since each scanner was operated only under one particular kVp and slice thickness, these variables could not be included in the statistical analysis (because they are therefore perfectly collinear with the type of scanner; see discussion below). Scan time and mA jointly determine dose, and hence only the latter has to be considered. Thus, for estimation purposes the only pertinent scanning parameter is dose. Since the two available measures of dose—peak and summation—rendered the same results in the regression analysis, only peak dose will be used throughout. As to the performance variables, each scan comprised six contrast "sectors", and Cohen calculated both nominal and two actual measures of contrast for each sector; the average of the three is the value used here. Threshold perceptibility, i.e., the diameter of the pattern of holes just visualized, was assessed by three different observers in each of these sectors, their average being the actual value used. Table II shows an example of the set of observations generated by each scan.

TABLE I. Description of data set.

Scanner	Number of scans	Number of observations	Dose range (rads)	Mean dose (rads)
AS&E	9	54	2.85-12.2	6.6
EMI	2	12	3.9	3.9
GE7800	10	60	0.42-6.45	2.16
GE8800	8	48	0.42-6.45	2.25
Pfizer	4	20 ^a	2.2-4.8	3.5
Searle	18	108	1-7.6	4.33
Total	51	302	0.42-12.2	3.9

^aObservations belonging to contrast sector 1 were omitted.

TABLE II. Example of observations from one scan. (Scanner: Pfizer—dose: 4.8 rads).

Sector	Contrast (%)	Perceptibility (mm)
1	0.25	11.20
2	0.53	8.67
3	0.83	4.07
4	1.23	2.48
5	2.22	1.77
6	3.04	1.47

B. Specifying the model

Hereafter the following notation will be used:

D : dose (in rads); $d = \log_e D$,

C : contrast (in %); $c = \log_e C$,

R : threshold perceptibility (in mm); $r = \log_e R$,

F : scanner-specific constant (or "scanner effects");

$f_i = \log_e F_i$.

Equation (1) can be rewritten now in a form suitable for estimation:

$$r = f_i - \beta_1 d - \beta_2 c + \epsilon, \quad (2)$$

where ϵ is an error or disturbance term (more on it later), and f_i , $i = 1, \dots, 6$, β_1 , and β_2 are the parameters to be estimated. To recall, the underlying theory predicts $\beta_1 = 1/3$ and $\beta_2 = 2/3$.

As already mentioned, in the few instances that an actual estimation was attempted (as in Ref. 1), separate bivariate regressions were used instead of a single multivariate one. There are two major problems with such a procedure: first, the f_i 's cannot be estimated consistently in such a way, and indeed they were not. Second, the rule is that *all* explanatory variables have to be included in the regression; leaving out any variable will bias the coefficients of the included variables, provided that the "left-out" and the included variables are statistically correlated ("bias" means that the obtained coefficients will be systematically lower, or higher, than the true coefficients). In the present case, leaving out dose, for example, will certainly be a source of bias, because different dose ranges were used for different scanners, i.e., the left-out dose variable is indeed correlated with the scanner-specific constants.

Now to the choice of the dependent, or "left-hand", variable: whereas in a mathematical equation it doesn't make any difference which variables are placed in the right- or left-hand side of it (i.e., $y = f(x)$ and $x = f^{-1}(y)$ are exactly equivalent, provided only that the inverse is well defined), that is not the case in the *statistical* specification of an equation. The key to the problem lies in the nature of the error term ϵ (which existence provides the rationale for the statistical estimation in the first place). In order to obtain consistent and unbiased estimates, ϵ has to fulfill several conditions, one of the most important being that it has to be uncorrelated with the explanatory variables. Therefore, the choice of left- and right-hand variables has to be such that this condition is ensured. In the present case the choice is straightforward. ϵ is the result of both the intrinsic randomness of the systems

as reflected in the received images, and of pure "errors of measurement", i.e., of random factors affecting the assessment of perceptibility by the different observers. Hence, it is clearly a component of perceptibility, but there is no reason to suspect that it will be correlated with dose, contrast or scanner type, which are the predetermined variables. The correct specification is thus as in Eq. (2), i.e., the one having perceptibility as the dependent variable. Cohen,¹ among others, took contrast to be the dependent variable instead, and it can be easily proven that that results in an *underestimate* of the coefficient of perceptibility, and hence in an *overestimate* of its inverse, the coefficient of contrast in an equation like Eq. (2).

C. Estimating the "scanner effects" (f_i)

Equation (2) assumes that, although the structural relationship between r , d , and c is the same for all scanners, they may differ from one another by a proportionality factor embodied in the f_i 's. In order to actually estimate them, one has to define "dummy variables" for the different scanners, and include them in the equation. A dummy variable assigns a value of 1 to the observations that correspond to what the variable is supposed to account for (in this case, belonging to a specific scanner), and zero to the others. When the equation includes an intercept and there are n scanners, only $n - 1$ dummies are to be used, the n th being already accounted for by the intercept. For convenience, the dummy variables will be denoted by the shorthand names of the scanners, and the intercept will stand for Pfizer. The equation now becomes

$$r = f_1 + \alpha_2 \text{AS\&E} + \alpha_3 \text{EMI} + \alpha_4 \text{GE78} + \alpha_5 \text{Searle} + \alpha_6 \text{GE88} - \beta_1 d - \beta_2 c + \epsilon, \quad (3)$$

and the "scanner effects" obtain as follows:

$$\begin{array}{ll} \text{Pfizer: } f_1, & \text{GE78: } f_4 = f_1 + \alpha_4, \\ \text{AS\&E: } f_2 = f_1 + \alpha_2, & \text{Searle: } f_5 = f_1 + \alpha_5, \\ \text{EMI: } f_3 = f_1 + \alpha_3, & \text{GE88: } f_6 = f_1 + \alpha_6. \end{array}$$

D. Functional form

As suggested above, it is implausible that the linear model (in the logs) actually holds over a wide range of values of the independent variables, because of the existence of absolute bounds both on perceptibility and on contrast. Moreover, even in between these bounds the behavior of a CT system is not uniform. For example, a distinction is often made between the "MTF dominated" (i.e., high-contrast) and the "noise dominated" (i.e., low-contrast) regions, the systems being assumed to be more responsive to contrast, and perhaps also to dose variations in the latter than in the former (Cohen and DiBianca² consider also an intermediate "transition region"). Although correct in themselves, these distinctions carry two undesirable features: They are *arbitrary* in setting the exact limits of each region, and they imply a sharply discontinuous or "jumpy" behavior, which is highly unlikely. These considerations suggest that the proper specification of the model requires a double asymptotical, *concave* function, that will hold throughout the entire range of the

TABLE II. Example of observations from one scan. (Scanner: Pfizer—dose: 4.8 rads).

Sector	Contrast (%)	Perceptibility (mm)
1	0.25	11.20
2	0.53	8.67
3	0.83	4.07
4	1.23	2.48
5	2.22	1.77
6	3.04	1.47

B. Specifying the model

Hereafter the following notation will be used:

D : dose (in rads); $d = \log_e D$,

C : contrast (in %); $c = \log_e C$,

R : threshold perceptibility (in mm); $r = \log_e R$,

F : scanner-specific constant (or "scanner effects");

$f_i = \log_e F_i$.

Equation (1) can be rewritten now in a form suitable for estimation:

$$r = f_i - \beta_1 d - \beta_2 c + \epsilon, \quad (2)$$

where ϵ is an error or disturbance term (more on it later), and $f_i, i = 1, \dots, 6, \beta_1$, and β_2 are the parameters to be estimated. To recall, the underlying theory predicts $\beta_1 = 1/3$ and $\beta_2 = 2/3$.

As already mentioned, in the few instances that an actual estimation was attempted (as in Ref. 1), separate bivariate regressions were used instead of a single multivariate one. There are two major problems with such a procedure: first, the f_i 's cannot be estimated consistently in such a way, and indeed they were not. Second, the rule is that *all* explanatory variables have to be included in the regression; leaving out any variable will bias the coefficients of the included variables, provided that the "left-out" and the included variables are statistically correlated ("bias" means that the obtained coefficients will be systematically lower, or higher, than the true coefficients). In the present case, leaving out dose, for example, will certainly be a source of bias, because different dose ranges were used for different scanners, i.e., the left-out dose variable is indeed correlated with the scanner-specific constants.

Now to the choice of the dependent, or "left-hand", variable: whereas in a mathematical equation it doesn't make any difference which variables are placed in the right- or left-hand side of it (i.e., $y = f(x)$ and $x = f^{-1}(y)$ are exactly equivalent, provided only that the inverse is well defined), that is not the case in the *statistical* specification of an equation. The key to the problem lies in the nature of the error term ϵ (which existence provides the rationale for the statistical estimation in the first place). In order to obtain consistent and unbiased estimates, ϵ has to fulfill several conditions, one of the most important being that it has to be uncorrelated with the explanatory variables. Therefore, the choice of left- and right-hand variables has to be such that this condition is ensured. In the present case the choice is straightforward. ϵ is the result of both the intrinsic randomness of the systems

as reflected in the received images, and of pure "errors of measurement", i.e., of random factors affecting the assessment of perceptibility by the different observers. Hence, it is clearly a component of perceptibility, but there is no reason to suspect that it will be correlated with dose, contrast or scanner type, which are the predetermined variables. The correct specification is thus as in Eq. (2), i.e., the one having perceptibility as the dependent variable. Cohen,¹ among others, took contrast to be the dependent variable instead, and it can be easily proven that that results in an *underestimate* of the coefficient of perceptibility, and hence in an *overestimate* of its inverse, the coefficient of contrast in an equation like Eq. (2).

C. Estimating the "scanner effects" (f_i)

Equation (2) assumes that, although the structural relationship between r, d , and c is the same for all scanners, they may differ from one another by a proportionality factor embodied in the f_i 's. In order to actually estimate them, one has to define "dummy variables" for the different scanners, and include them in the equation. A dummy variable assigns a value of 1 to the observations that correspond to what the variable is supposed to account for (in this case, belonging to a specific scanner), and zero to the others. When the equation includes an intercept and there are n scanners, only $n - 1$ dummies are to be used, the n th being already accounted for by the intercept. For convenience, the dummy variables will be denoted by the shorthand names of the scanners, and the intercept will stand for Pfizer. The equation now becomes

$$r = f_1 + \alpha_2 \text{AS\&E} + \alpha_3 \text{EMI} + \alpha_4 \text{GE78} + \alpha_5 \text{Searle} + \alpha_6 \text{GE88} - \beta_1 d - \beta_2 c + \epsilon, \quad (3)$$

and the "scanner effects" obtain as follows:

$$\begin{aligned} \text{Pfizer: } & f_1, & \text{GE78: } & f_4 = f_1 + \alpha_4, \\ \text{AS\&E: } & f_2 = f_1 + \alpha_2, & \text{Searle: } & f_5 = f_1 + \alpha_5, \\ \text{EMI: } & f_3 = f_1 + \alpha_3, & \text{GE88: } & f_6 = f_1 + \alpha_6. \end{aligned}$$

D. Functional form

As suggested above, it is implausible that the linear model (in the logs) actually holds over a wide range of values of the independent variables, because of the existence of absolute bounds both on perceptibility and on contrast. Moreover, even in between these bounds the behavior of a CT system is not uniform. For example, a distinction is often made between the "MTF dominated" (i.e., high-contrast) and the "noise dominated" (i.e., low-contrast) regions, the systems being assumed to be more responsive to contrast, and perhaps also to dose variations in the latter than in the former (Cohen and DiBianca² consider also an intermediate "transition region"). Although correct in themselves, these distinctions carry two undesirable features: They are *arbitrary* in setting the exact limits of each region, and they imply a sharply discontinuous or "jumpy" behavior, which is highly unlikely. These considerations suggest that the proper specification of the model requires a double asymptotical, *concave* function, that will hold throughout the entire range of the

variables. A plausible candidate displaying the required properties will be

$$r = \bar{r}_i + \gamma_1(\bar{d} - d) + \gamma_2/(c - \bar{c}_i) + \epsilon, \quad (4)$$

(for $c > \bar{c}$ and $d < \bar{d}$),

where \bar{c}_i represents "floor noise" of system i , \bar{d} the effective upper dose limit (in the sense that beyond it changes in dose will not affect perceptibility) and \bar{r}_i the approximate limiting resolution of the system,⁶ i.e., as $c \rightarrow \infty$ and $d \rightarrow \bar{d}$, $r \rightarrow \bar{r}_i$.

However, estimating a function like Eq. (4) poses serious problems. It has to be done with nonlinear methods, which are obviously more complex and expensive and, more important, it requires data covering a wide range of values, including observations at the neighborhood of the asymptotes. Such data are not yet available, and hence a less demanding specification has to be sought. Two possibilities are at hand: a piecewise linear form, and adding a quadratic term for c in Eq. (2).

The piecewise function is a formalization of the above-mentioned distinction between contrast regions. Following Cohen,¹ the low-contrast region is taken to be that for which $C \leq 1\%$ (i.e., $c \leq 0$), hence including contrast sectors 1-4 (see Table II), whereas the transition region includes contrast sectors 5 and 6. In order to allow for different contrast coefficients for each region within a single regression, an interaction variable has to be created, as follows: define a dummy variable p so that $p = 1$ for observations belonging to sectors 5-6, and $p = 0$ otherwise; the interaction variable is then $K = p \cdot c$, and the equation to be estimated becomes

$$r = f_1 + \sum_{i=2}^6 \alpha_i S_i - \beta_1 d - \beta'_2 c + \beta_3 K + \epsilon, \quad (5)$$

where S_i is the scanner i . The slope for the low-contrast region is then $-\beta'_2$, and that for the transition region ($-\beta'_2 + \beta_3$). Note that this procedure is very different from Cohen's. He simply deleted the observations of the transition region, thus losing valuable information and rendering a less efficient estimation.

The second option consists of adding the quadratic term c^2 :

$$r = f_1 + \sum_{i=2}^6 \alpha_i S_i - \beta_1 d - \beta''_2 c + \beta'_3 c^2 + \epsilon. \quad (6)$$

In order to compare this with alternative formulations, it is convenient to make use of the concept of the *elasticity* η_{yx} of a variable y with respect to x , measuring the percentage change in y resulting from a one percentage change in x , and defined as

$$\eta_{yx} \equiv \frac{\partial y}{\partial x} \frac{x}{y} = \frac{\partial \log y}{\partial \log x}. \quad (7)$$

It follows that in Eq. (5) the coefficient β_1 represents the elasticity of perceptibility with respect to dose, and β'_2 and $(\beta'_2 - \beta_3)$ the contrast elasticities in the low-contrast and transition regions, respectively. In Eq. (6), though, the contrast elasticity is not a constant but a function of c ,

$$-\eta_{RC} = \beta''_2 - 2\beta'_3 c. \quad (8)$$

The two alternatives are compared diagrammatically in Fig. 1. Note that the coefficient β''_2 in the quadratic model is the

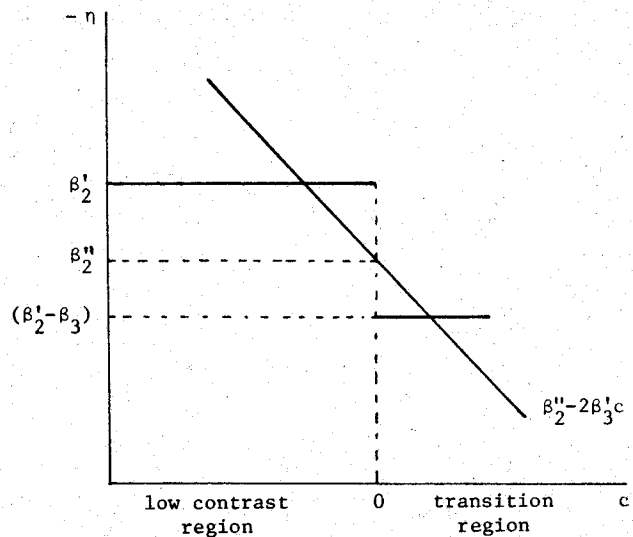


FIG. 1. Contrast elasticities in the piece-wise linear and in the quadratic models (hypothetical example).

elasticity at $C = 1\%$ ($c = 0$), the contrast value that lies in between the geometric and the arithmetic means of the actual range of contrasts studied (0.8 and 1.2, respectively), setting the limit between the low-contrast and transition regions. Thus, β''_2 has in this case a very helpful interpretation, namely, it represents the *average response* of the systems over the studied interval.

Although the quadratic specification is conceptually superior to the piecewise linear in that it avoids the arbitrariness of the limits between regions and the discontinuous response, it makes little difference in the present case which one is actually used in estimating the model. The reason is that the available range of contrast is too narrow (to recall, $0.17\% < C < 3\%$) to allow for the fine distinction between a concave and a piecewise linear function. Both make equally good *local* approximations, and will fit the data just as well. This underlines the fact that, at this stage (i.e., after having arrived at the correct specification), the real constraint is the availability of data. In order to be more precise, and be able to obtain a *full* characterization of performance, one would need data covering the behavior of the systems over their entire range.

IV. EMPIRICAL RESULTS

I present first the estimated regressions, and then analyze separately the results for contrast, dose, and scanner effects. Both "pooled" (or "common") and individual (or scanner-specific) regressions were estimated, the former comprising all the data, and the latter estimating a different regression for each scanner. As expected, the coefficients obtained in the quadratic and in the linear models are very much alike (for convenience, the piecewise linear model will be called hereafter just linear). Thus, and in view of the conceptual advantages of the quadratic, only its results will be used in the discussion of the dose coefficient and scanner effects. All coefficients except for one (to be noted below) were found to be statistically significant at the 0.99 level, and therefore there was no need to report t ratios. The pooled regressions

TABLE III. Individual regressions—quadratic model.

Scanner	Constant	Coefficients of			R^2
		d	c	c^2	
1. Pfizer	1.510	-0.219	-0.913	0.185	0.961
2. AS&E	1.315	-0.2	-0.687	0.292	0.966
3. EMI	1.140	a	-0.589	0.359	0.976
4. GE78	1.221	-0.191	-0.673	0.194	0.951
5. Searle	1.125	-0.132	-0.600	0.152	0.968
6. GE88	1.142	-0.226	-0.805	0.186	0.985

^a Could not be estimated because only one dose was used.

are shown below, and the individual regressions in Tables III and IV. The pooled regression-quadratic model is

$$r = 1.418 - 0.054 \text{AS\&E} - 0.178 \text{EMI} - 0.18 \text{GE78} \\ - 0.26 \text{Searle} - 0.268 \text{GE88} - 0.176d \\ - 0.676c + 0.16c^2, \quad R^2 = 0.95. \quad (9)$$

(* : significant at the 0.8 level only).

Pooled regression, linear model:

$$r = 1.365 - 0.081 \text{AS\&E} - 0.194 \text{EMI} - 0.199 \text{GE78} \\ - 0.274 \text{Searle} - 0.281 \text{GE88} - 0.177d \\ - 0.949c + 0.517K, \quad R^2 = 0.95. \quad (10)$$

A. Contrast elasticities

As shown in Table V, the contrast elasticity varies substantially along the range of contrast studied, thus supporting the presumption of concavity: it goes, in the pooled regression, from 0.3 at the highest contrast value ($C = 3\%$), to 1.2 at the lowest ($C = 0.2\%$), averaging 0.43 over the transition region and 0.95 over the noise region. But the most important finding is the following:

The contrast coefficient of 2/3 predicted in the theoretical model holds *exactly* at a 1% contrast, and can be interpreted to represent the *average response* of a CT system within a low to medium contrast range.

This aspect of the theory is thus confirmed in a restricted sense. However, given that the function is nonlinear, one should warn against using the value of 2/3 to characterize

TABLE IV. Individual regressions—linear model.

Scanner	Constant	Coefficients of			R^2
		d	c	I	
1. Pfizer	1.489	-0.219	-1.107	0.416	0.970
2. AS&E	1.224	-0.207	-1.120	0.828	0.970
3. EMI	1.197	...	-0.920	0.721	0.925
4. GE78	1.137	-0.190	-0.997	0.617	0.955
5. Searle	1.056	-0.132	-0.873	0.511	0.970
6. GE88	1.104	-0.223	-1.061	0.487	0.980

the behavior of CT systems outside the range considered here. Moreover, this result refers to the average response of the six scanners: if interested in a particular scanner and/or a particular subset of the range, the exact value should be calculated using the pertinent coefficients as detailed in Table V.

Another finding of interest is that the behavior of the different scanners is much more uniform in the noise region than in the transition region (as measured by the SD/mean of the coefficients). That is, all systems improve their perceptibility to a similar extent when increasing contrast at low-contrast levels, but some, like the EMI 1005 and the AS&E almost exhaust their capabilities when going up to 2%–3% contrast, whereas the others still have margins for further improvements (these being higher for Pfizer and GE88 than for Searle and GE78).

B. The dose coefficient

Table III shows that, putting Searle aside for the moment, the dose coefficients of the four scanners are very similar, averaging 0.209 (with a SD/mean value of only 0.08). This result differs sharply from the value of 1/3 predicted by the theoretical model, and may have far-reaching implications for the actual dosage used in CT. It means that substantial variations in dose have little effect on image quality (e.g., doubling dose improves spatial resolution by only 20%) and hence for many applications the dose delivered could be reduced significantly without impairing much the information content of the received images. To put it differently, in order

TABLE V. Contrast elasticities.

	Linear model			Quadratic model	
	"noise" region $0.2\% \leq C < 1\%$	"transition" region $1\% < C \leq 3\%$	Average $C \sim 1\%$	Average $C = 1\%$	Extreme values $C = 0.2\%$ $C = 3\%$
Pooled	0.95	0.43	0.69	0.676	1.19 0.32
Pfizer	1.11	0.69	0.90	0.91	1.51 0.50
GE88	1.06	0.57	0.82	0.80	1.40 0.39
AS&E	1.12	0.29	0.71	0.69	1.63 0.05
GE78	1.00	0.38	0.69	0.67	1.29 0.24
Searle	0.87	0.36	0.62	0.60	1.09 0.27
EMI	0.92	0.20	0.56	0.59	1.75 0 ^b
SD/mean ^a	0.10	0.44	0.18	0.17	0.16 0.58(0.8) ^c

^a A measure of the variability of results across scanners: the standard deviation over the mean of the coefficients of the six scanners.

^b The strong concavity of the EMI regression results, in fact, in a negative value here, which is meaningless; thus, for all practical purposes it should be considered to be zero.

^c It is 0.58 for the 5 scanners excluding EMI, and 0.8 if EMI is included and its value taken to be zero.

to optimize in the use of CT, it is of prime importance to know the exact magnitudes of the trade-offs built in the systems. A dose coefficient of 1/5 means that for any optimization setting, the solution (i.e., the optimal dose) will be only 60% of that that would result if a coefficient of 1/3 was assumed instead. Haaga *et al.*¹³ arrived at a similar qualitative conclusion and moreover, the same quantitative result can be obtained using the data presented in Table I of their article; regressing their noise measure (RMSD) on dose (in mAs) gives

$$\log(\text{RMSD}) = 2.82 - 0.305 d, \quad r^2 = 0.88. \quad (11)$$

In order to compare this result with the above, the dose coefficient has to be multiplied by 2/3 (i.e., the average coefficient of contrast, or noise, predicted by the theoretical model and confirmed, with the said caveats, by my findings) rendering a value of 0.203. This almost exact identity of results is particularly significant, given that they used a very different research design, and gathered the data from a scanner not studied here (the Ohio Nuclear 2010).

Now to Searle: the substantially lower dose coefficient obtained for it suggested two hypotheses, namely, that the dose response is related to the fact that the system was operated with three different sampling rates (yielding 360, 540, and 1080 "views" per scan), and/or that the perceptibility-dose relationship is not linear but concave. Both were confirmed, but the second is more revealing. If running different regressions for each sampling rate, one obtains a dose coefficient of 0.1 for scans of 360 and 540 views, and a coefficient of 0.16 for 1.080 views. Thus, lower sampling rates reduce indeed the dose response of the system, and the 0.132 coefficient represents not a uniform behavior but only the resulting average. However, the highest coefficient (0.16) is still significantly less than the benchmark of 0.20. As to concavity, a regression of Searle's data including a quadratic term for dose rendered the following results:

$$r = 1.19 - 0.6c + 0.15c^2 - 0.31d + 0.08d^2, \quad (12)$$

$$R^2 = 0.97.$$

The coefficient of d^2 being significant at the 0.996 level, hence strongly supporting the hypothesis. Applying the formula in Eq. (7) to calculate the dose *elasticities* at different dose levels (those actually used with Searle's) gives

Dose:	1	1.9	3.8	7.6
Elasticity:	-0.31	-0.21	-0.1	~0

Thus, the dose response declines rapidly as dose increases (0.132 being the average over the range), and it coincides with the benchmark only at around 2 rad. What remains to be explained is therefore not Searle's low coefficient in itself, but the reasons for its sharp drop (of the other scanners, only the GE7800 shows signs of concavity, but to a lesser extent).

C. Indices of performance

Before turning to the estimates of the f_i 's, it is important to understand their precise meaning and establish how these may legitimately be used. If it were the case that the same equation held for all scanners, i.e., that $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$,

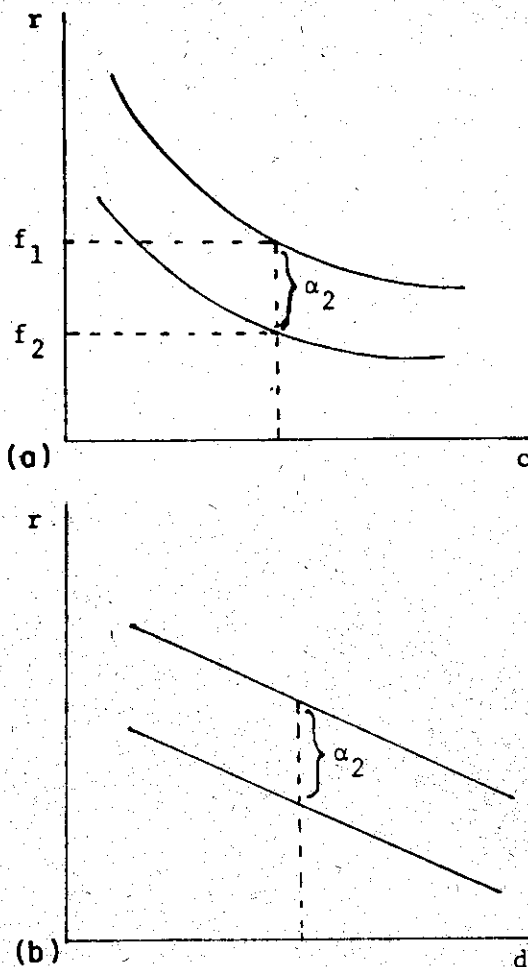


FIG. 2. Diagrammatical representation of "figures of merit". Case 1: The same equation holds for the two systems being compared. (a) In r, c space (assuming $d = 0$); (b) in r, d space.

and $\beta_{3i} = \beta_3$ for all $i = 1, 2, \dots, 6$, then the f_i 's would indeed be a well-defined summary parameter of the image quality of each system.¹¹ In other words, if we take a particular scanner, say f_1 , to be the baseline, the differences $\alpha_i \equiv (f_i - f_1)$ would be the appropriate figures to be used in the construction of indices of image quality. This is shown in Figs. 2(a) and 2(b), representing typical "slices" of the three-dimensional function relating r to c and d . Clearly, the pooled and the individual regressions will produce in this case the same estimates of the f_i 's.

However, if some of the coefficients differ among scanners (that is, if the differences are statistically significant, as these turned out to be in the present case),¹² then the f_i 's estimated in the individual regressions will not coincide with those of the pooled regression and moreover, they will no longer have the above, straightforward meaning. Figures 3(a) and 3(b) typify this case. α_2 measures the vertical distance between the curves for systems 1 and 2, *only* at the points where the independent variables equal zero. However, these points are as good as any others and, given that the difference between the systems varies widely over the range, the α 's so measured cannot be taken as indicative of the "overall" advantage of one scanner upon the other.

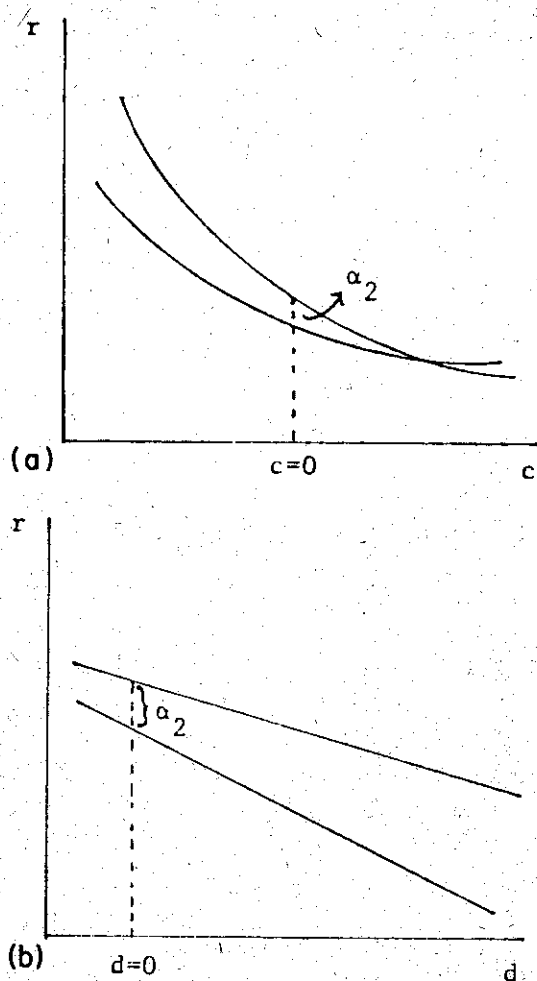


FIG. 3. Diagrammatical representation of "figures of merit". Case 2: Each system renders a different equation. (a) In r, c space; (b) in r, d space.

Ideally, what is needed to overcome this comparability problem is a valuation function $V = V(r, c, d)$, relating the performance dimensions to diagnostic usefulness. In principle, this function could be derived from systematic clinical studies in the line of Robbins *et al.*¹⁴ Although it is hard to imagine that a precise correspondence between, say, low- and high-contrast resolution and diagnostic value would ever be established, it is reasonable to expect that some qualitative notions regarding these correspondences will be developed. The role of such a function would be to assign weights to every point (or region) over the range, thus enabling the construction of a summary index of performance consisting of the weighted average of the distances between the curves. Lacking as of now a better alternative, it is reasonable to assume a *uniform* V function, i.e., one that assigns equal weights to all points over the range.

This is, of course, only a tentative assumption (and so will be the derived results) to be modified if and when pertinent evidence suggesting a different specification becomes available. But it has two distinctive advantages. First, it serves as a convenient reference point, in the sense that it is easy to assess how the indices would change as a result of changing the weighting scheme in a particular direction. Second, and more important, the scanner specific constants correspond-

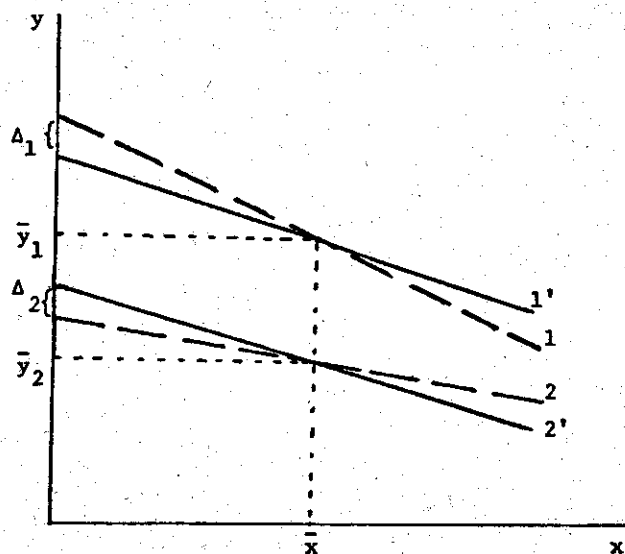


FIG. 4. The effect of "pooling" in regression analysis. The broken lines represent the individual regressions: $y = f_i + \beta_i x, i = 1, 2$. The solid lines correspond to the pooled regression: $y = f_i^0 + \beta^0 x$.

ing to that assumption (i.e., those representing the *average* differences between scanners) are in fact those obtained from the pooled regression (with some caveats), and are therefore readily available. This can be easily demonstrated as follows: the average "height" of a function is by definition the mean value of its dependent variable \bar{y} ; given that the constant term f in a regression is calculated simply as $f = \bar{y} - \bar{x}$, the fitted curve must go through the point (\bar{y}, \bar{x}) . Thus, all that is needed is to calculate the distance between the curves of different scanners at the point of the mean values of the independent variables. But that is precisely what the difference between the constants in the pooled regression do, as can be seen in Fig. 4: the effect of pooling is to impose the same slope coefficient β^0 on both scanners, thus rotating the individual curves by $(\beta_i - \beta^0)$ at \bar{x} . The constant terms of the pooled regression are therefore

$$f_i^0 = \bar{y}_i - \beta^0 \bar{x} = f_i + \Delta_i, \quad \Delta_i = (\beta_i - \beta^0), \quad (13)$$

and the "advantage" of system 2 over system 1 is

$$\alpha_2 = f_2^0 - f_1^0 = \bar{y}_2 - \bar{y}_1. \quad (14)$$

Applying Eq. (13) to the present case (\bar{d} and \bar{c} represent mean values)

$$f_i^0 = f_i - (\beta_{1i} - \beta_1^0)\bar{d} - (\beta_{2i} - \beta_2^0)\bar{c} + (\beta_{3i} - \beta_3^0)\bar{c}^2, \quad (15)$$

which is intuitively appealing: the higher the dose response of a system, and the lower its contrast elasticity (all measured at their respective averages), the smaller f_i^0 will be and hence the better the system is.

A crucial assumption in the above argument is that all scanners have the same mean values of the independent variables. Otherwise, the f_i 's will capture also differences in these values, and therefore the α_i 's will be measuring not only relative performance, but differences in the ranges studied as well. The range of contrast is quite uniform across scanners: it averages 1.2% with a SD/mean of only 0.12, but the dose range varies widely (see Table I). In order to achieve

TABLE VI. Indices of performance.

Scanner	(1) α_i	(2) f_i	(3) F_i	(4) q_i	(5) q'_i
1. Pfizer	0	1.390	4.015	1	1
2. AS&E	-0.055	1.336	3.804	0.94	0.76
3. EMI	-0.175	1.215	3.370	0.84	(0.42)*
4. GE78	-0.181	1.209	3.350	0.83	0.40
5. Searle	-0.260	1.130	3.096	0.77	(0.27)*
6. GE88	-0.281	1.109	3.031	0.75	0.25

*These figures were calculated assuming a dose coefficient of 1/5; but, to recall, no such coefficient could be estimated for EMI for lack of data and the one for Searle was found to be 0.132.

a reasonable degree of uniformity without losing too much information, those scans taken with very low and very high doses were excluded: one from AS&E taken with $D = 12.2$, and three from each of the GE scanners, taken with $D < 1$. The common range is now $1 < D < 7.6$, with a mean value of 3.9 rad and $SD/mean = 0.2$. The resulting regression is

$$r = 1.39 - 0.055AS\&E - 0.175EMI - 0.181GE78 - 0.26Searle - 0.281GE88 - 0.16d - 0.66c + 0.16c^2, \\ R^2 = 0.952. \quad (16)$$

Note that the slope coefficients are almost identical to those in Eq. (9), the regression including the whole data set, which is important so as to make the results here consistent with those in previous sections. The scanner-specific constants are also very similar although not identical to the ones in Eq. (9), but in view of the fact that slight variations in them do make a difference for purposes of comparing performance, Eq. (16) should indeed be used rather than Eq. (9). The results are presented in Table VI: the F_i 's correspond to what is known as "figures of merit", and are calculated simply as $F_i = \exp(f_i^0)$. The question now is how to construct on the basis of the F_i 's an index of imaging performance, i.e., a unit-free parameter that will enable to compare easily among scanners. An obvious candidate will be (taking again Pfizer to serve as the baseline)

$$q_i = \frac{F_i}{F_1} = \frac{\exp(f_i^0 + \alpha_i)}{\exp(f_1^0)} = \exp(\alpha_i). \quad (17)$$

The interpretation of q_i is straightforward. On average, the perceptibility of scanner i is $(q_i \times 100)\%$ that of Pfizer, i.e., scanner i is $[(1 - q_i) \times 100]\%$ better than Pfizer in terms of perceptibility. Column 4 in Table VI shows the values of q_i for each scanner, which appear to cluster in three groups: Pfizer and AS&E, EMI and GE78, and Searle and GE88. There is a 13% difference between the second and the first groups, and a 7% between the third and the second, whereas the differences within each group are very small, and only marginally significant from a statistical viewpoint.

An index of performance is obviously not unique, but depends upon the choice of the variable with respect to which the scanners are being compared: q_i does it in terms of perceptibility, but it could be done in terms of contrast or dose as well. The latter is of particular interest, and the question to which the appropriate index should respond is; how much dose does scanner i require, relative to the base, in order to

achieve a certain perceptibility and given a certain contrast? This is equivalent to making D the left-hand variable in Eq. (1), and remembering that the dose coefficient for four scanners (i.e., except for EMI and Searle) was found to be 0.2, all that has to be done is to raise the F_i 's to the power of the inverse of that coefficient, i.e., to the 5th power. The new index, q'_i to be called dose efficiency, would then be $q'_i = q_i^5$. As shown in Table VI, column 5, the differences between scanners are obviously magnified when these refer to dose requirements. Whereas, say, the GE88 produces images with 25% better perceptibility, on the average, than Pfizer's, it requires 75% less dose to attain the same image quality.

It is important to note and stress the differences between these indices and measures of performance proposed elsewhere in the literature. As stated above, Southon⁹ calculated "figures of merit" F for each mode of operation of each scanner, and did so by substituting the observed values for the corresponding variables in a model such as Eq. (1), taken as given. I have already pointed out that such procedure is *a priori* incorrect (i.e., the F_i 's have to be estimated along with the other parameters in the context of a complete model, not calculated by substitution) and my results indicate that, in fact, Eq. (1) is not the correct model. Moreover, Southon's findings themselves show that his procedure is unwarranted. The F_i 's are supposed to be by definition a constant for each scanner, but these were found in Ref. 9 to vary widely with mode of operation for a given system (in some cases by as much as a factor of 8, i.e., many times more than what could be accounted for by the mere statistical fluctuation of the data). This just indicates that the F_i 's are picking up the effect of variables that were not included in the model (i.e., "left-out" variables) and that vary for each scanner (e.g., kVp in the EMI5005), in addition to the effect of using the wrong coefficients for the included variables (e.g., 0.5 instead of 0.3 for dose).

Cohen's measure of a perceptibility-dose-factor (PDF) in Ref. 2 can be criticized along similar lines, although the problem there is less apparent because he calculates the PDFs for only one mode of operation for each scanner.

V. CONCLUDING REMARKS

The bulk of the efforts in previous studies of CT performance has been devoted to the collection of data on various variables of interest and related problems of measurement. It has been the purpose here to show that the statistical methods of analyzing such data, in the context of a methodologically consistent framework, matter as much. The finding regarding the magnitude of the dose coefficient underlines the practical importance of this claim, and so do the estimated indices of image quality, especially when compared to previous measures.

Although the study has been confined to a contrast-detail analysis of CT scanners, I insist that the approach put forward here (i.e., the use of multivariate regression analysis), should be applied to any quantitative study of imaging performance (of any modality) that involves more than two dimensions, and/or attempts to perform inter or intrasystems comparisons.

Finally, it is worth noting that the analysis here can help

clarify the controversy generated by an editorial in the *American Journal of Roentgenology* by Hasegawa *et al.*¹⁵ [see the ensuing correspondence in *Am. J. Roentgenol.* **135**, 1310 (1980)]. The main arguments raised there, if correctly interpreted, refer in fact to the fallacies that may result from estimating *partial* models (e.g., "leaving-out" dose, or kVp if it were the case that the latter influences the measure of contrast). This paper hopefully demonstrates that the conclusion in Ref. 15 regarding the inadequacy of "contrast-detail" studies in general is totally unwarranted, and stems from the same sort of methodological confusion.

ACKNOWLEDGMENTS

The data for this study were provided by Dr. Gerald Cohen, for which I am most grateful. I also wish to thank Philip Judy, Zvi Griliches, Alan Garber, John Bound, a reviewer from BRH, and two anonymous referees for helpful comments on an earlier draft. I am thankful for the financial support that was provided for this work by the NSF through Grant No. SOC 79-04279.

^{a)} Present address: Tel-Aviv University, Department of Economics, Ramat Aviv, Tel-Aviv 69978, Israel.

^{b)} These methods are widely in use in other disciplines, primarily in behavioral and biological sciences.

^{c)} Obviously, this is relevant only when one of the goals is indeed the assessment of the *relative* performance of different models of a given imaging modality. It is not as yet clear whether such comparisons are possible and/

or meaningful *across* different imaging modalities; if that proves to be feasible (at least for some modalities and some dimensions of imaging performance), then the statistical methods proposed here would be all the more necessary.

^{d)} Zatz (Ref. 11), for example, claims a factor of 10 for otherwise identical functions, but it is hard to evaluate his conclusions given the lack of detailed numerical results or data.

^{e)} "Approximate" because $c < 100$ and hence the term $\gamma_2(c - \bar{c}_i)$ doesn't reach zero.

^{f)} I shall use hereafter "image quality" to indicate only those aspects of the imaging performance of a scanner encompassed in the "contrast-detail" framework of analysis. Needless to say, other factors such as artifactual behavior, uniformity, linearity, etc., should also be taken into account for a more complete description of a system.

^{g)} This was established by performing the "F test" for pooling, which was clearly rejected.

¹G. Cohen, *J. Comput. Assist. Tomogr.* **3**, 197 (1979).

²G. Cohen and F. A. DiBianca, *J. Comput. Assist. Tomogr.* **3**, 189 (1979).

³L. K. Wagner, G. Cohen, W. H. Wong, and S. R. Amtey, *Med. Phys.* **8**, 24 (1981).

⁴R. G. Gould, B. Belanger, H. I. Goldberg, and A. Moss, *Radiology* **137**, 783 (1980).

⁵S. W. Smith and H. Lopez, *Med. Phys.* **9**, 4 (1981).

⁶R. A. Brooks and G. DiChiro, *Phys. Med. Biol.* **21**, 689 (1976).

⁷G. N. Hounsfield, *Am. J. Roentgenol.* **127**, 3 (1976).

⁸D. A. Bassano, Ch. C. Chamberlain, J. M. Mozley, and S. A. Kieffer, *Radiology* **123**, 455 (1977).

⁹F. C. Southon, *Med. Phys.* **8**, 62 (1981).

¹⁰R. F. Wagner, D. G. Brown, and M. S. Pastel, *Med. Phys.* **6**, 83 (1979).

¹¹L. M. Zatz, *J. Comput. Assist. Tomogr.* **2**, 336 (1978).

¹²D. R. White, R. D. Speller, and P. M. Taylor, *Br. J. Radiol.* **54**, 221 (1981).

¹³J. R. Haaga, F. Miraldi, W. MacIntyre, J. P. LiPuma, P. J. Bryan, and E. Wiesen, *Radiology* **138**, 449 (1981).

¹⁴A. H. Robbins, R. D. Pugatch, S. G. Gerzof, L. J. Faling, W. C. Johnson, R. Spira, and D. R. Gale, *Radiology* **137**, 719 (1980).

¹⁵B. H. Hasegawa, R. K. Cacak, J. A. Mulvaney, and W. R. Hendee, *Am. J. Roentgenol.* **135**, 199 (1980).