

A Graphical Analysis of the Interrelationships among Waterborne Asbestos, Digestive System Cancer and Population Density

by Michael E. Tarter,* Robert C. Cooper,*† and William R. Freeman‡

Five statistical procedures were used to partial the correlation between waterborne asbestos and digestive site cancer for the putative effects of population density. These include: analysis based on a data subset with roughly homogeneous population density; standard residual analysis (partial correlation); conditional probability integral transformation; analysis based upon ranked data, and use of logarithmic transformation.

Nonparametric regression graphical techniques are applied to examine the nature or shape of the asbestos-cancer dose-response curve. Evidence is presented that suggests that there is considerable difference between analyses involving nonhigh-density tracts and non-San Francisco tracts. Evidence is also presented that the modal-type nonparametric regression curve forks or bifurcates when adjustment is made for population density.

Introduction

In 1979, a study was completed by Cooper et al. (1), which dealt with certain health effects of asbestos in the domestic water supplies of five California counties. Several publications (2,3) and even a lawsuit involving the State of Connecticut and asbestos pipe producers have resulted in part from the tentative findings of the Cooper study that "a statistical association between asbestos fiber content in drinking water and the incidence of certain cancers in the San Francisco Bay Area has been shown."

In testimony presented during the Connecticut hearings, one witness suggested that the results of the Cooper-Kanarek et al. study were suspect due to lack of adjustment for, or partialling of, the

population density variate. This paper describes a follow-up investigation that was undertaken to deal with the questions involving the population density putative intervening variate.

The material presented here is not reported upon in order of scientific significance, but instead the simplest methodological approach is described first and subsequent results are organized in terms of the complexity of the statistical methodology used. For example, an investigation in which the original Cooper et al. (1) study is repeated on a subset of low and moderately populated census tracts is described. By far the most interesting and dramatic reported here it is that presented at the end of this paper, which relies on a complex type of nonparametric regression (one based on a locus of conditional density modes). It demonstrates that the morbidity density may be bimodal for high asbestos exposure and unimodal for low asbestos exposure census tracts.

Exclusion of the Highly Dense Tracts

The estimated distribution of census tract sub-grouping population density variate Y is highly

*Department of Biomedical and Environmental Health Sciences, University of California at Berkeley, Berkeley, CA 94720

†Sanitary Engineering and Environmental Health Research Laboratory, University of California at Berkeley, Berkeley, CA 94720.

‡Graphical Biometry Laboratory, University of California at Berkeley, Berkeley, CA 94720.

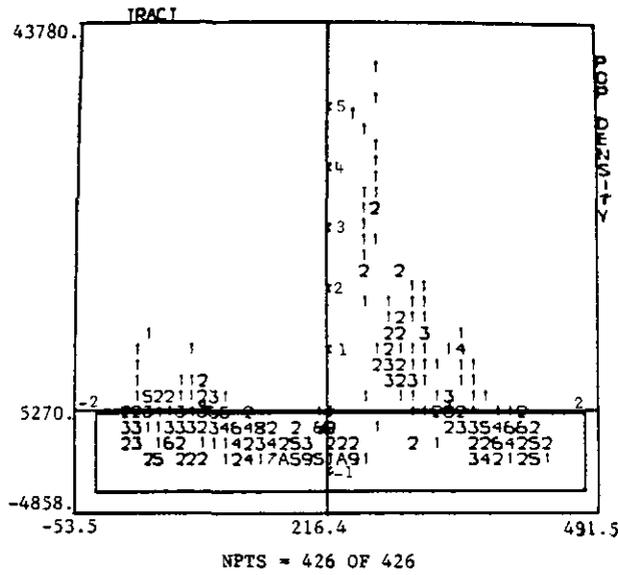


FIGURE 1. Population density-census tract frequency diagram.

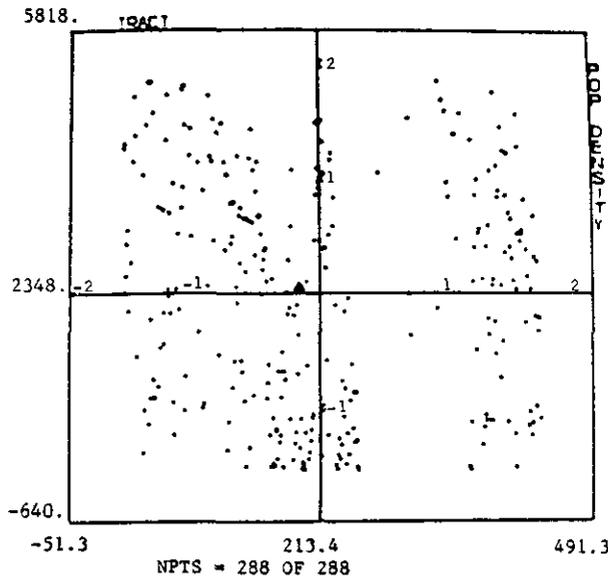


FIGURE 2. Low to moderate density scatter diagram.

skewed. Figure 1 displays the frequency distribution of Y and the independent variate tract number code X . To avoid confusion (e.g., between 11 coincident points and 2 nearby single points, each represented by the symbol '1'), a frequency of 10 is represented by the character A, 11 by B, 12 by C, and so on up to 35 (Z) and more (*). As an indication of skewness, note that while most census tracts contain less than the mean $E(Y) \equiv 5270$ individuals per square mile, several census

tracts are at least six times more densely populated than the mean. Relative to tracts where Y takes on a value > 5270 , tracts with $Y < 5270$ form an almost homogeneous subgrouping. Furthermore, tract number (the X variate of Fig. 1) is a rough measure of geographic location. Hence, the Bay Area population seems to consist of the following three components: (1) a highly dense component lying slightly to the right of the mean tract number $X = E(X) \equiv 216$ line; (2) a moderately dense component below component 1 but above the $Y = E(Y) \equiv 5270$ line; and finally (3) a large group of tracts with moderate to low densities, shown in Figure 1 enclosed by an editing rectangle.

When the region enclosed by the editing rectangle is enlarged and a scatter diagram of the region is displayed as in Figure 2, there appears to be a moderately homogeneous spread of Tract POP DENSITY points in most portions of the display not immediately to the right of the $X = 216$ line.

Studies of the variate log population density (described below) suggest that there may be two distinct classifications within what appears by eye to be one homogeneous grouping of low to moderate density census tracts. Nevertheless, as a preliminary analysis, it seems reasonable to repeat the basic Cooper et al. (1) study with the use of the low to moderate population density tracts.

There were 288 tracts with low to moderate population density. Figure 3 is a frequency dia-

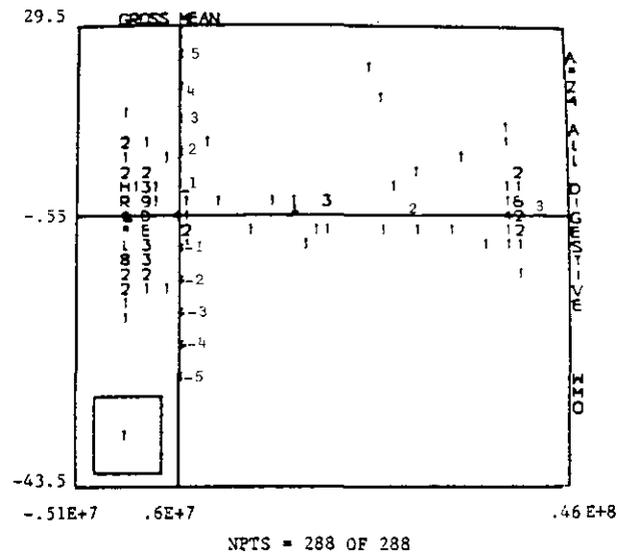


FIGURE 3. Low to moderate density asbestos-excess cancer frequency diagram.

gram obtained by plotting the variate asbestos exposure in fibers/liter (indicated as "gross mean" in the figure) on the *x*-axis and the variate observed minus expected (i.e., excess) white male all digestive site per tract cancer on the *y*-axis (1). One census tract whose coordinates are shown enclosed in an editing rectangle appears to have either an extremely low observed cancer morbidity or an extremely high expected cancer morbidity. The position of the point that represents this tract implies that, were this tract to be removed, the Pearsonian correlation between the *X* and *Y* variates would be reduced. Nevertheless, despite the removal of this point, and much more importantly the removal of all high-density tracts from consideration, the correlation between asbestos exposure and excess cancer morbidity variates is 0.194. Even for a one-sided test, the α level for the null hypothesis of no association would have to be smaller than 0.005 for one to accept the null hypothesis in this case. Thus one can be reasonably sure that the exclusion of highly populated tracts from the original study of Cooper et al. (1) would not change the study's principal finding of a significant relationship between waterborne asbestos exposure and digestive tract cancer. It might be noted that the correlation between asbestos exposure and excess morbidity for the entire data set of low, medium, and high-density tracts is 0.241.

Regression Studies

In a previous study (3), nonparametric regression procedures were applied in order to study the

shape of the dose-response curve associated with the asbestos exposure-cancer excess morbidity relationship. In this section it will be demonstrated that the previously observed linearity of the dose-response curve seems to persist for low to moderate exposure levels and low to moderately populated tracts.

Following the approach used in Tarter (3) as a preliminary to the analysis of the 287 low to moderate density census tracts described by Figure 4, the margins between the leftmost and rightmost data points and the endpoints of the interval of estimation were reduced. This operation was performed for the purpose of taking advantage of the periodic nature of Fourier series nonparametric representation. The effects of this margin reduction are shown in Figure 5, which can be thought of as an enlargement of Figure 4.

The conventional least-squares fitted regression line obtained from the 287 low to moderate density points is shown in Figure 6, as are the 95% confidence bands associated with this line. These bands enclose a region within which, provided certain basic assumptions are satisfied (4), one is 95% sure that the true dose-response curve is contained. The fact that no horizontal, i.e., null relationship, line is contained within the confidence bands is the graphical equivalent of the dose-response significant relationship described in the previous section.

Unlike the parametric regression line shown in Figure 6, nonparametric regression procedures do not rely on the initial assumption that the dose-response curve conforms to any particular functional relationship e.g., line. However, the flexi-

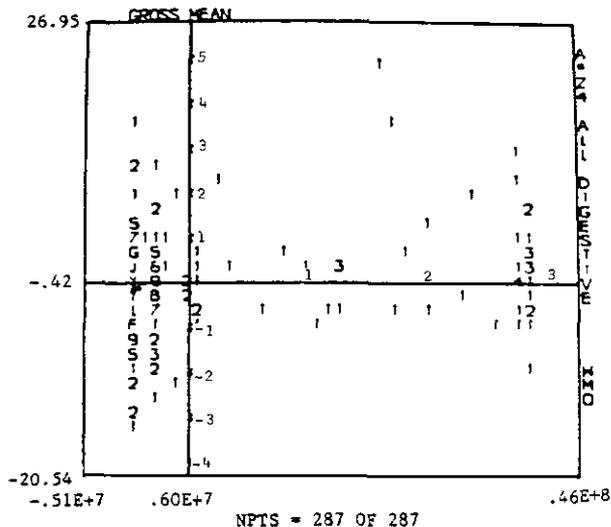


FIGURE 4. Frequency diagram after removal of outliers.

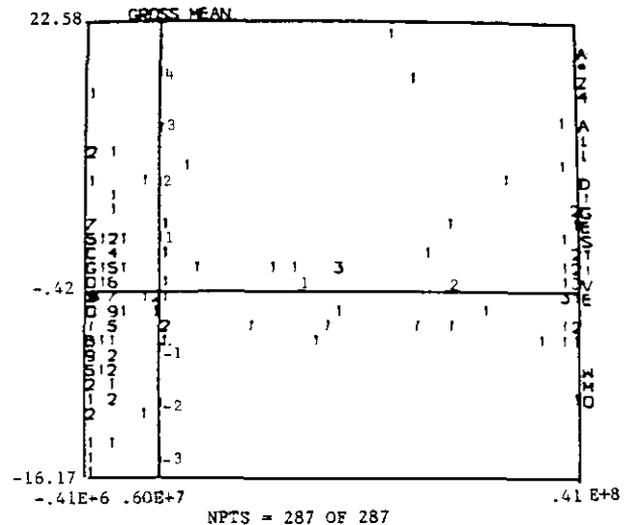


FIGURE 5. Margin-reduced frequency diagram.

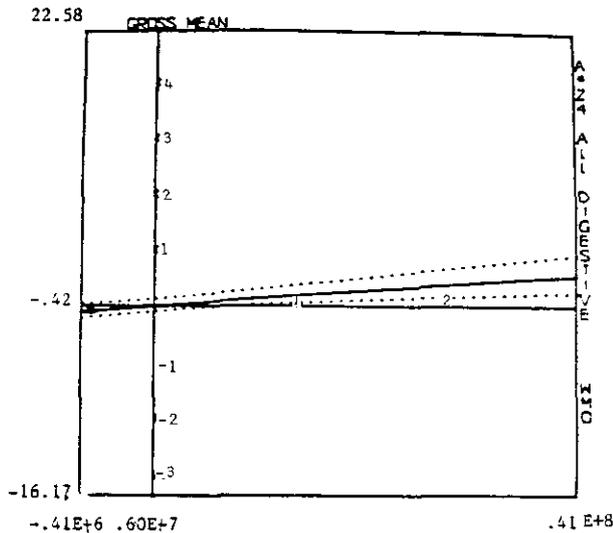


FIGURE 6. Regression line with confidence bands.

bility inherent in nonparametric methodology makes it extremely difficult to obtain useful estimates within subregions where data points are sparsely distributed. Approaches that can be used to augment conventional procedures and effectively deal with sparse-rich data mixtures are described by Breiman et al. (5), Tarter (6), and Bean and Tsokos (7).

Figure 7 is a contour diagram that illustrates the uneven distribution of asbestos exposure-excess cancer morbidity data. To obtain Figure 7 a generalized three-dimensional histogram was constructed above the gross mean asbestos exposure and excess digestive cancer sample space. This estimated frequency surface was then sectioned parallel to the sample space to produce the contours, i.e., terraces, isopleths, shown in Figure 7. The three outermost contours correspond to loci of points whose estimated relative frequency is exactly 30% of the modal frequency, i.e., the most common single combination of exposure and excess morbidity.

Naturally, the nonparametric regression curve shown in Figure 8, which was obtained from the same estimated bivariate frequency used to obtain Figure 7, exhibits a highly irregular pattern. The procedure used to obtain this curve has been described in Tarter (8).

To increase the resolving power of the nonparametric estimation procedure, a technique was employed that increases the second moment of all density components in the x -direction but that leaves all other distributional characteristics unchanged. This procedure is a variant of a spectral

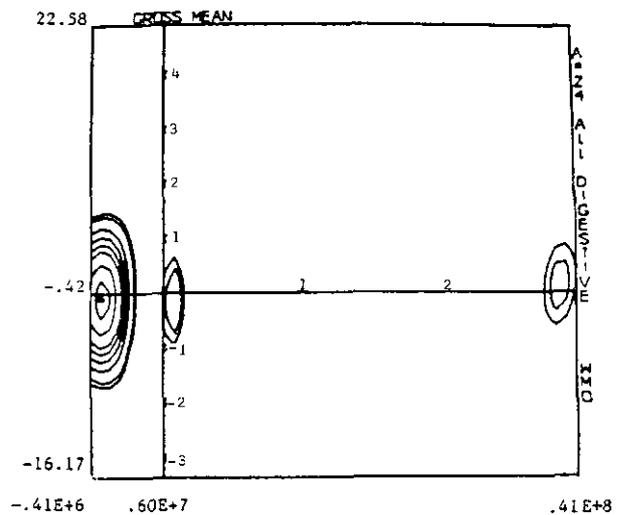


FIGURE 7. Contour diagram showing sparse-rich data distribution.

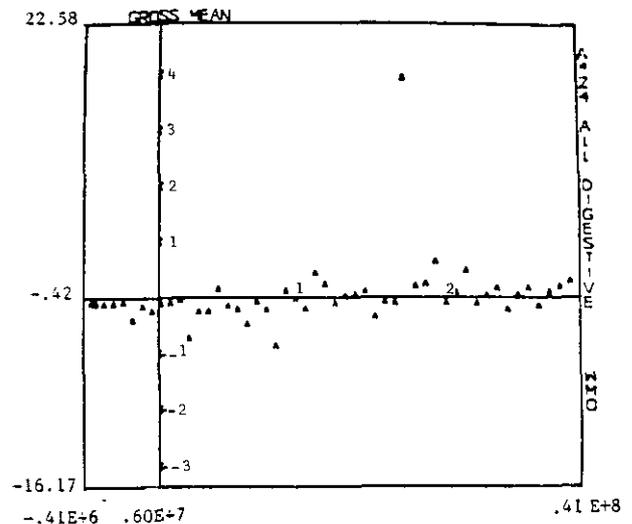


FIGURE 8. Nonparametric regression before smoothing.

analysis technique described by Medgyessy (9) and generalized to two dimensions by Stanat (10) and Tarter and Silvers (11). Figure 9 is the analog of Figure 7 produced by smoothing in the x -direction, while Figure 10 is the analog to Figure 8.

As depicted by the indicator or pointer line, the nonparametric regression curve seems to reach a plateau or horizontal component at coordinates $X = 27,500,000$ fibers/L and $Y = 0.51$ excess cases over a 6-yr period per census tract. In other words, when only the low to medium density tracts are considered, there is no statistical evidence that a change of exposures to any value in

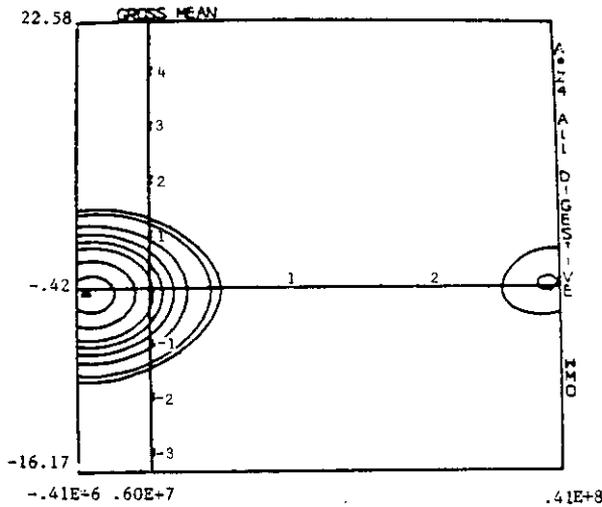


FIGURE 9. Equiprobability contours after smoothing.

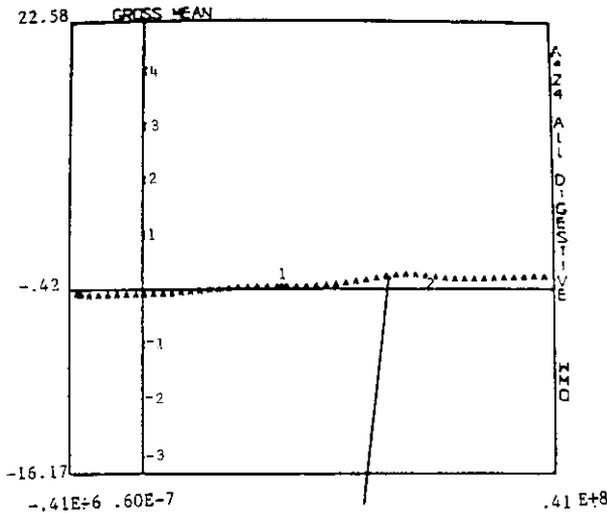


FIGURE 10. Nonparametric regression after smoothing.

excess of 27,500,000 fibers/L will increase the estimated dose-response curve to more than 0.51 excess cases. This finding is of course far from surprising, since there are so few high-exposure tracts among the 287 tracts with low to medium population density.

By using procedures described by Tarter (3), benchmark data confirmation procedures were applied to check whether the statistical methods used to obtain Figure 8 could have influenced the ramp as opposed to step i.e., diagonal as opposed to horizontal shape of the nonparametric regression curve. These procedures tend to confirm that not only is the relationship between waterborne asbestos level and excess cancer statistically signifi-

cant for the low to moderate population density subsample, but there exists concomitant variation (12) between these two variates up to the point $X = 27,500,000$, $Y = 0.51$. When nonparametric regression procedures other than that used to obtain Figure 8 were applied to these data (13), the same slight but discernible concomitant increase in Y with an increase in X less than 27,500,000 fibers/L was observed.

Residual and Partialling Studies

The most straightforward and elementary procedure usually used to remove the effects of a single underlying or nuisance variable is of course residual linear regression or its near equivalent, partial linear correlation. One can view this procedure in terms of a process that involves the following two basic steps. (1) One or both variables to be partialled are used as response or dependent Y variates, and the variate Z whose effect is to be removed is fitted (usually by least-squares methods) to Y . Most commonly a regression slope $\hat{\beta}$ is obtained such that Y and $\hat{\beta}Z$ have maximum correlation. Up to a constant, usually denoted by a , $\hat{\beta}Z$ "explains" as much of the variation of Y as is possible by any linear combination $a + \hat{\beta}Z$. (2) The residual $Y - \hat{\beta}Z$ is used to replace the response variate Y , and in some instances the same procedure involving a second value of $\hat{\beta}$ is used to find the residual of principal independent variable X and nuisance independent variable Z . For the problem at hand, Y represents excess cancer morbidity, X asbestos exposure level, and Z population density. The variate $Y - \hat{\beta}Z$ is sometimes interpreted as that part of Y that has not been "explained" by the variate Z (14).

For the 426 original census tracts of Cooper et al. (1), Figure 11 shows the result of applying the above residual regression technique. To better interpret this figure, a code has been used in our GRAFSTAT program (15-18), which on the right-hand vertical column specifies B*Z2 A*Z1 ALL DIGESTIVE WMO. To interpret this graph legend, note that the basic Y variate was the ALL DIGESTIVE site white male observed (hence WMO) per census tract number of cases per 6-yr period. The prefix A*Z1 designates that the A (linear or additive) transformation has been applied to the Z1 variate ALL DIGESTIVE WME, where the last "E" designates that it was the expected number of cancer cases that was substituted for the observed cases to form, after subtraction from its WMO counterpart, the basic Y variate, excess morbidity.

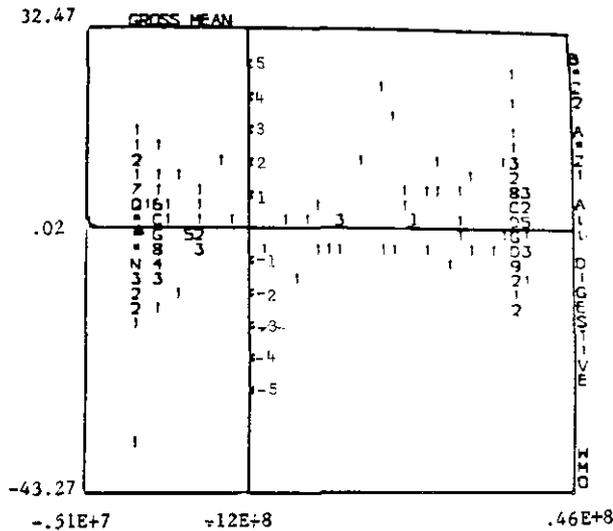


FIGURE 11. Standard partialled frequency diagram.

The outer prefix B*Z2 designates that after the variate excess morbidity had been formed the least-squares residual (transformation B) was used to obtain that part of the A*Z1 ALL DIGESTIVE variate that was not linearly explained by the Z2 variate, population density.

A comparison of Figures 11 and 3 shows a very close correspondence between the B transformed complete data set and the untransformed low to medium density data subset at all but the rightmost column of data points. The same lower left-hand corner putative outlier is present in these two figures, and the correlation between the two variates is $r = 0.171$, which for a sample of size $n = 426$ is significant at the $\alpha = 0.01$ level.

It should be mentioned that we have repeated all the experiments described in this paper with the X variate gross mean asbestos exposure adjusted for population density by the same technique used to adjust the Y variate, morbidity. For ease of interpretation, experiments involving a partialled or residual Y variate and an unpartialled (original) X variate will be presented here.

Data with the single lower leftmost putative outlier removed from the file were used to obtain Figure 12 and an associated correlation of $r = 0.171$.

Conditional Probability Integral Partialling

The partialling or residual approach described in the preceding section has the advantages of being well studied and easily interpreted pro-

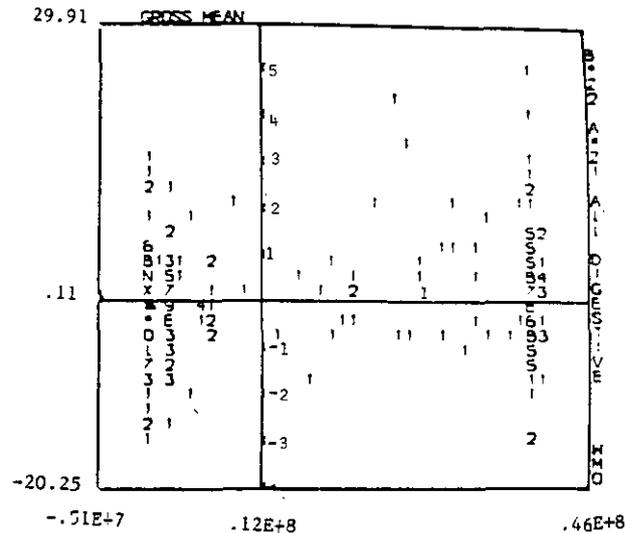


FIGURE 12. Partialled frequency diagram after removal of outliers.

vided certain initial specifications or assumptions are satisfied. The so-called "problem of specification" (19) underlies a considerable literature usually denoted by the terms nonparametric, model-free, or curve estimation. Hence, it seems very reasonable to consider nonparametric alternatives to simple linear-residual analysis.

As will be described later in detail, the variate population density has a highly nonnormal and skewed distribution. Hence, the first of the wide variety of model-free approaches applied involved a technique known as a conditional probability integral transformation (20, 21) whose effectiveness is minimally affected by distributional specification.

Unfortunately the price paid for generality is a certain difficulty in interpretation. The variate I*Z2 A*Z1 ALL DIGESTIVE WMO, where the letter I designates that the conditional probability integral information (CPIT), has been applied to remove the effect of population density, has a correlation of $r = 0.133$ with the asbestos exposure variate, as shown in Figure 13. It is interesting that the positive correlation persists despite the drastic nature of the CPIT transform. However, it is too early in the study of the CPIT method to accurately assess the statistical significance of the correlation in this case.

Parametric-Nonparametric Hybrids

Two approaches were used to take into account the skewed nature of the density variate and yet

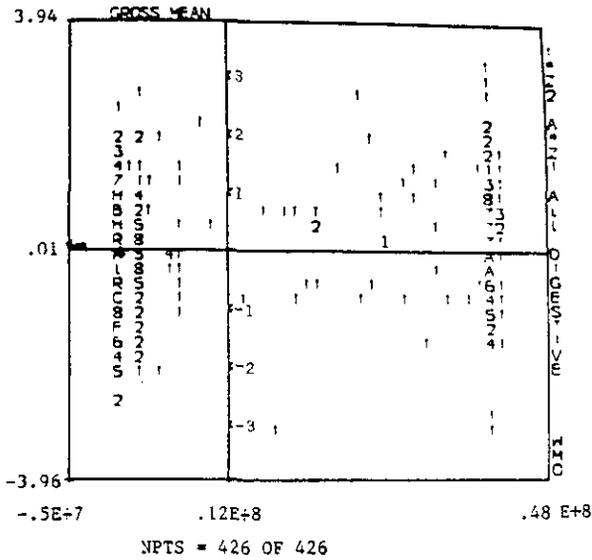


FIGURE 13. Conditional probability integral partialled frequency diagram.

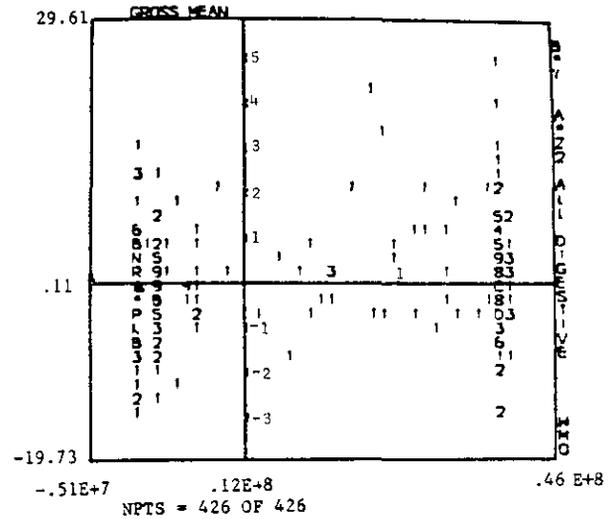


FIGURE 15. Partialled and log transformed data.

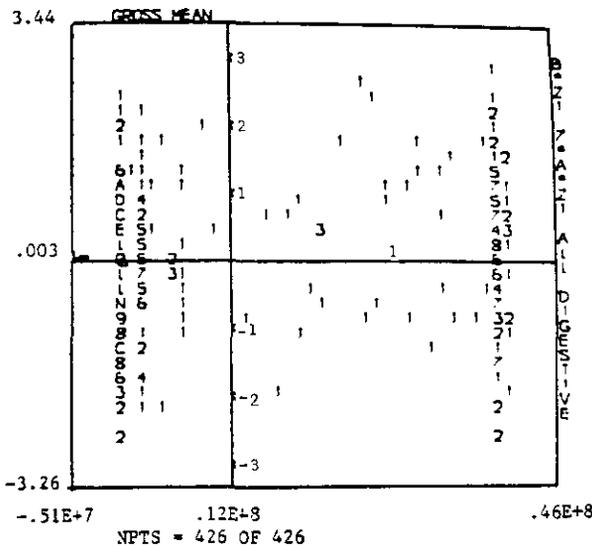


FIGURE 14. Partialled and ranked data.

obtain a result more easily interpreted in terms of statistical significance than the result described in the previous section. The first involved a commonly used nonparametric approach based on the rank transformation (22), and the second the logarithmic transformation (3).

The number 7 shown in the Y axis label of Figure 14 designates the seventh GRAFSTAT data transformation option, normalized rank or score (4). For a fixed value of nuisance variable Z, e.g., population density, this option assures that

the order between any two transformed values matches that of any two untransformed values, and that the distribution of the transformed score is approximately standard Gaussian which, among other things, guarantees that the variate's distribution is no longer skewed. The correlation in this case is $r = 0.143$.

It should be pointed out that although the distribution of both the 7*A*Z1 ALL DIGESTIVE WMO variate and the 7*POP DENSITY variate can now be treated as if they were Gaussian, the joint distribution of Y and X need not be bivariate normal. Nevertheless, as described by Tarter and Kowalski (21), the use of the normalized score transform does seem to be useful in a wide variety of contexts with the proviso, as suggested by the Figure 14 Y variate prefix B*Z1, that the simple least-squares difference between the normalized variates is sufficient to remove all density effect.

To obtain Figure 15, the variate log POP DENSITY was used in place of the ranked variate 7*POP DENSITY. The similarity between Figure 11 and Figure 10 and that between the corresponding correlations of $r = 0.157$ and $r = 0.143$ suggests that in this case the correlation is highly robust with respect to changes in the X and Y variate individually. The difference between the above correlation and the correlation associated with the CPIT I transformation suggests that the relationship between the X and Y variates may be highly complex. One possible cause of this complexity could be the existence of disjoint distributional subcomponents for either the morbidity or population density variate.

Univariate Investigation of the Population Density Variate

Figure 16 is a histogram constructed from the 426 census tracts. The variate whose density is estimated by this histogram is the logarithm of the per-tract population density. Figure 17 is a nonparametric kernel-Fourier series estimator (23), which suggests that the gap between the leftmost and second-leftmost mode of the histogram in Figure 16 is not an artifact traceable to sampling variation. This brings into question the

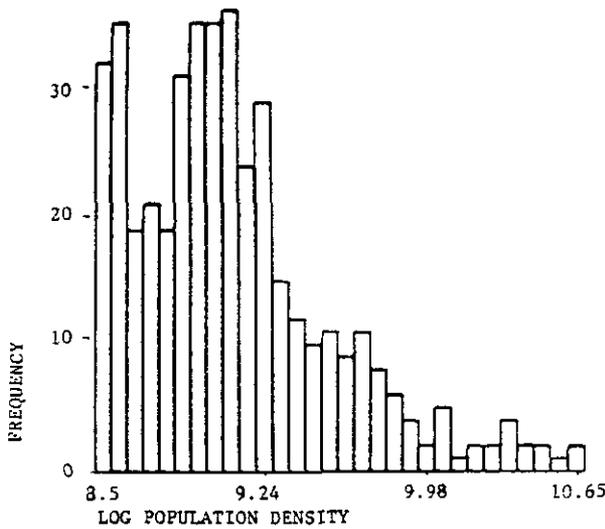


FIGURE 16. Log population density histogram.

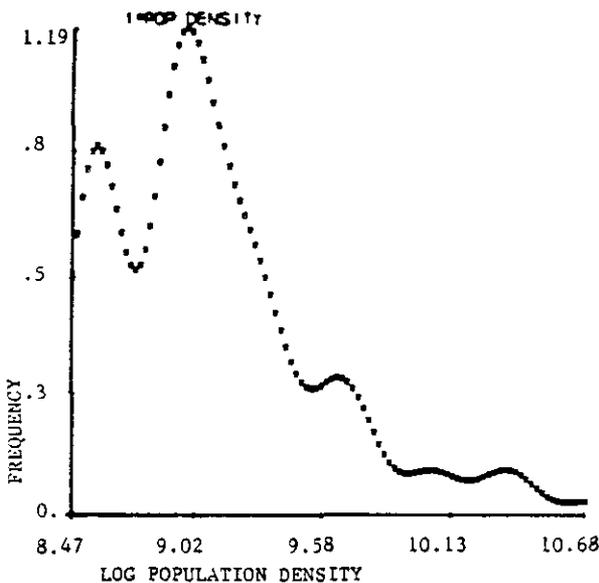


FIGURE 17. Nonparametrically estimated log population frequency curve.

low to moderate population density homogeneity that underlies the analysis.

Study of Non-San Francisco Tracts

Because of the complexity of the population density variate, it was decided to repeat the investigation described above after excluding tracts within the city limits of San Francisco. The relationship between the basic asbestos exposure X variate and the observed per (non-San Francisco) tract white male all digestive organ morbidity is described by Figure 18. The correlation between these two variates is $r = 0.136$. (Here sample size $n = 323$, i.e., there were 323 non-San Francisco tracts.)

Once the Z_1 variate, expected morbidity, was subtracted from the observed morbidity, Figure 19 was obtained. Now, however, the correlation between X and Y has dropped to the nonstatistically significant value of $r = 0.05$. Thus, one can be reasonably sure that the significant relationship between asbestos and cancer is largely attributable to those low to moderately populated tracts located in San Francisco. Tentative findings have already been published involving several San Francisco tracts (24) which suggests that there may be data collection or spurious correlation artifacts associated with a particular San Francisco subregion.

To carefully study the non-San Francisco tracts, the lower left outlier was again removed as shown

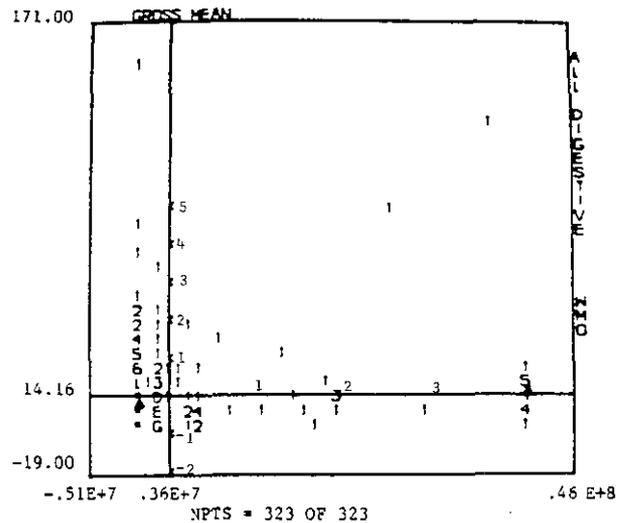


FIGURE 18. Non-San Francisco observed cancer-asbestos exposure frequency diagram.

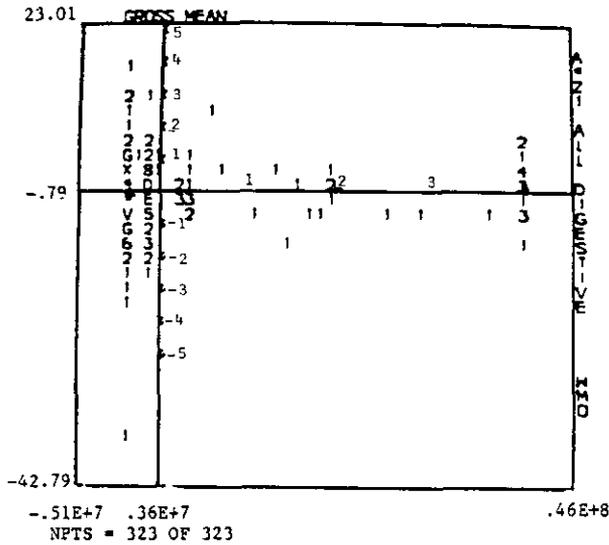


FIGURE 19. Non-San Francisco excess cancer-asbestos exposure frequency diagram.

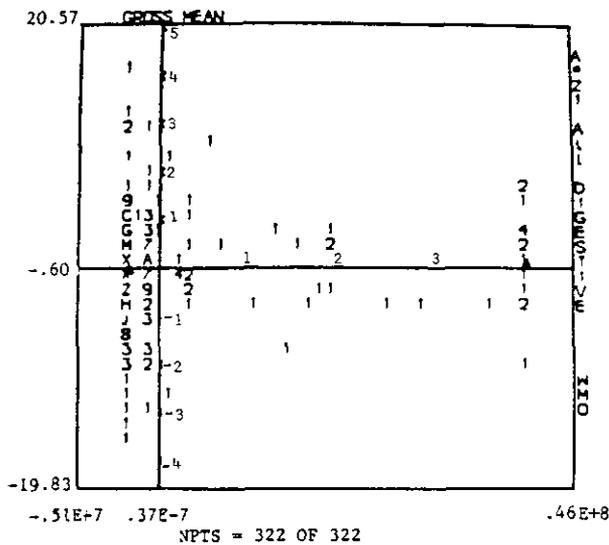


FIGURE 20. Frequency diagram after removal of outliers.

in Figure 20 and the tracts were divided into two groups, the first containing tracts with less than 3.6 million fibers/liter exposure and the second containing the remaining moderate to high exposure tracts (3.7 million is the mean exposure associated with the non-San Francisco tracts.)

When modern curve estimation procedures were applied to estimate the excess morbidity frequencies for these two data subsets, Figure 21 was obtained. Despite the fact that the methods used were completely automated and model-free

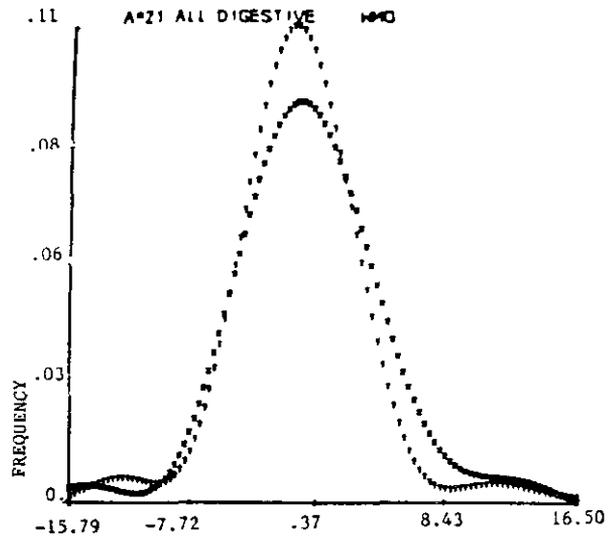


FIGURE 21. Estimated excess cancer frequency curves for high (Y) and low (X) exposure subpopulations.

and the high exposure excess morbidity density was estimated from only 45 observations (since most high exposure tracts lay within the boundaries of San Francisco), Figure 21 shows two curves that not only have almost exactly the same mean but the same variance and bell-like shape. The curve traced with the symbol X corresponds to the low-exposure, while the curve traced with the symbol Y describes the high-exposure group. When the cumulatives corresponding to Figure 21 densities were constructed, the maximum difference between these curves was found to be 0.083. The $\alpha = 0.05$ critical value for the Kolmogoroff-Smirnov test statistic for the sample sizes involved is very close to 0.21, which is many magnitudes greater than the estimated cumulative difference. Hence, if there is a causal rather than spurious association between waterborne asbestos exposure and digestive cancer morbidity, the existence of this relationship using Bay Area data can only be ascertained by careful study of differences between the San Francisco subpopulation and the subpopulations of other locations.

The Within-San Francisco Study

If one performs a study opposite to that described in the last section and restricts one's attention *only* to San Francisco tracts, Figure 22 is obtained. The observed correlation in this case is negative ($r = -0.114$) although not significant, since there are relatively few San Francisco tracts ($n = 99$).

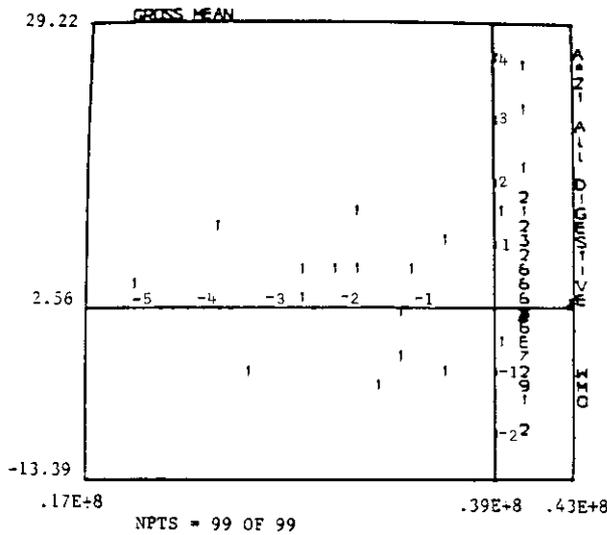


FIGURE 22. Exclusively San Francisco tracts.

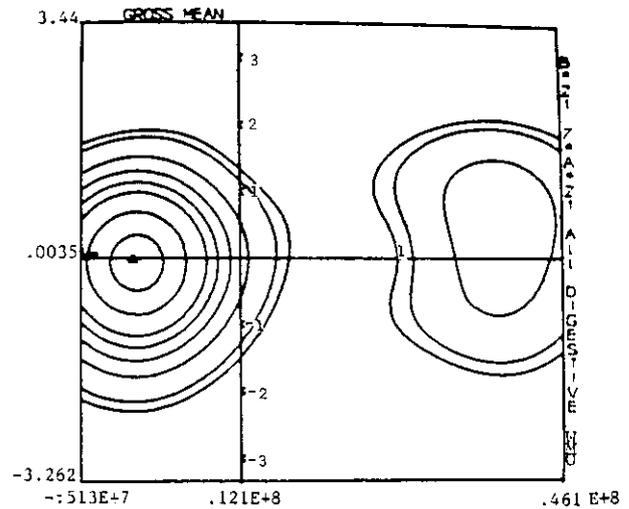


FIGURE 24. Equiprobability contours using ranked data.

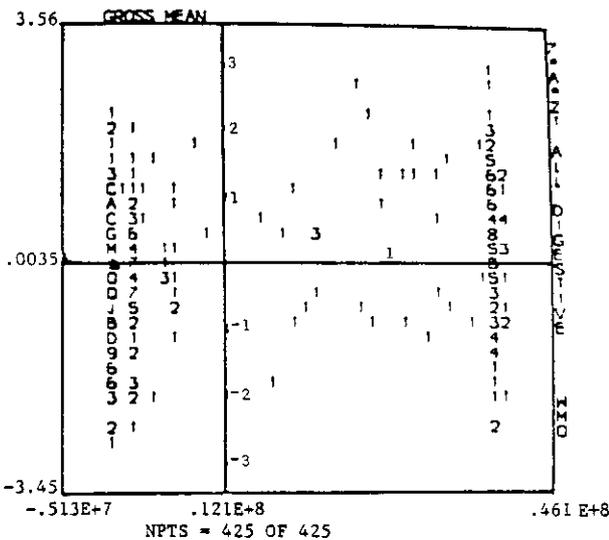


FIGURE 23. Use of ranking transformation after morbidity variate construction.

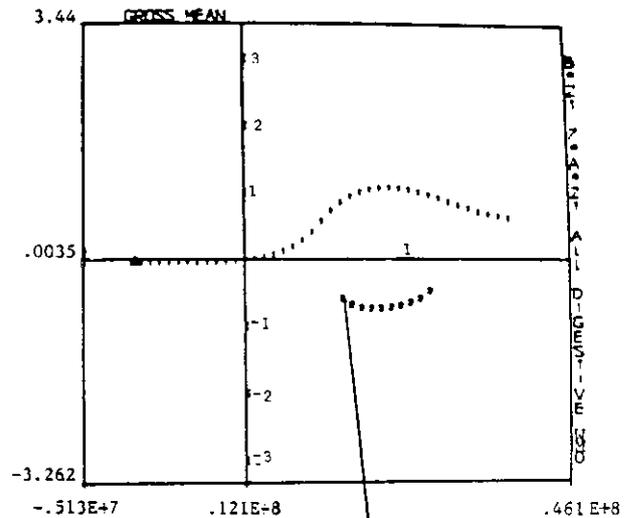


FIGURE 25. Nonparametric regression using ranked data.

Further Evidence for High-Exposure Distributional Subcomponents

One study has provided very suggestive evidence that the high asbestos exposure subpopulation contains two disjoint subcomponents. To obtain Figure 23, ranking transformation 7 was applied after the excess cancer morbidity variate was constructed. The use of transformation 7 to some extent corrects the blurring induced by the possibly highly complex interrelationship be-

tween observed and expected morbidity. Now when the data shown in Figure 23 are parialled, i.e., the residual of Z1 variate ranked density is constructed and when the smoothing technique described above is applied, the contours shown in Figure 24 was obtained. Much more importantly, when a nonparametric regression based on density modes is executed, Figure 25 is generated. As noted by the pointer shown in this figure, the high-exposure subpopulation itself seems to be made up of two distinct subpopulations for asbestos exposure levels above 20 million fibers/L.

Since almost all high-exposure tracts were lo-

cated in San Francisco, it seems that the relationship between waterborne asbestos exposure and excess digestive cancer morbidity may be traceable to a greater proportion of tracts associated with code 1 as shown in Figure 25 than tracts associated with code 2, i.e., the bimodal density of morbidity has a higher mode associated with high than with low morbidity as shown in Figure 25, where the number 1 is used to represent the higher and number 2 the lower mode. Hypothetically, suppose that there exist two types of census tracts, type 1 (higher mode) and type 2 (lower mode), the former with a greater and the latter with a lesser probability of an individual contracting digestive cancer. It appears from Figure 25 that the high asbestos exposure subgroup seems to have a greater preponderance of type 1 individuals than does the low to moderate exposure subpopulation. Of course, as suggested in the previous two sections, this finding may not have anything to do with a causal connection between asbestos and cancer. It may simply be a coincidence that type 1 tracts tend to be clustered in that part of the Bay Area that happens to have high exposure levels.

Despite the above caution, it should be pointed out that the rather striking nature of Figure 25 was only uncovered after a partialling technique was applied to remove the effect of population density. Thus, rather than being an artifact traceable to the underlying effects of population density, the relationship between asbestos levels and excess cancer morbidity was brought more clearly into focus once population density was taken into account.

The authors would like to thank Seemin Qayum and Joe Rodrique for their technical assistance. This research was supported by grant 1 RO1 CA 28142-01A2 from the National Cancer Institute, National Institutes of Health.

Preparation of this manuscript was supported by contract 68033130 from the U.S. Environmental Protection Agency.

The research described in this paper has been peer and administratively reviewed by the U.S. Environmental Protection Agency and approved for presentation and publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

- Cooper, R. C., Tarter, M. E., Kanarek, M. S., Murchio, J. K., Conforti, P. M., Jackson, L. A., Callard, R., and Lysmer, D. Asbestos in Domestic Water Supplies in Five California Counties. EHS Publication No. 79-1, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, 1979.
- Kanarek, M. S., Conforti, P. M., Jackson, L. A., Cooper, R. C., and Murchio, J. C. Asbestos in drinking water and cancer incidence in the Bay Area. *Am. J. Epidemiol.* 112: 54-72 (1980).
- Tarter, M. E. Pattern recognition in the context of an asbestos cancer threshold study. In: *Proceedings of the Thirteenth Annual Symposium on the Interface: Computer Science and Statistics* (W. F. Eddy, Ed.), Springer-Verlag, New York, 1981, pp. 105-110.
- Dixon, W. J., and Massey, F. J. *Introduction to Statistical Analysis*. McGraw-Hill, New York, 3rd ed., 1969.
- Breiman, L., Meisel, W., and Purcell, E. Variable kernel estimates of multivariate densities. *Technometrics* 19: 135-144 (1977).
- Tarter, M. E. Biocomputational methodology: an adjunct to theory and applications. *Biometrics* 35: 9-24 (1979).
- Bean, S., and Tsokos, C. P. Bandwidth selection procedures for kernel density estimates. *Commun. Statist. Theory Methods* 11: 1045-1069 (1982).
- Tarter, M. E. Trigonometric maximum likelihood estimation and application to the analysis of incomplete survival information. *J. Am. Statist. Assoc.* 74: 132-139 (1979).
- Medgyessy, P. *Decomposition of Superpositions of Distribution Functions*. Hungarian Academy of Sciences Publishing House, Budapest, 1961.
- Stanat, D. F. Unsupervised learning of mixtures of probability functions. In: *Pattern Recognition* (L. N. Kanal, Ed.), Thompson, Washington, DC, 1968, pp. 357-389.
- Tarter, M. E., and Silvers, A. Implementation and applications of bivariate Gaussian mixture decomposition. *J. Am. Statist. Assoc.* 70: 47-55 (1975).
- Mill, J. S. *A System of Logic, Ratiocinative and Inductive*. Longmans, Green, Reader, and Dyer, London, 8th ed., 1872, pp. 464-466.
- Fisz, M. *Probability Theory and Mathematical Statistics*. John Wiley, New York, 3rd ed., 1963, pp. 91-96.
- Croxtan, F. E., and Cowden, D. *Applied General Statistics*. Prentice-Hall, New York, 1940, p. 743.
- Tarter, M. E., Rigsbee, E. O., and Wong, J. T. Interactive editing of biomedical data. *Computer Programs Biomed.* 6: 117-123 (1976).
- Tarter, M. E., Marshall, J. S., Lum, S. B., Rigsbee, E. O., and Wong, J. T. A description of the Berkeley Graphical Biometry Project. In: *Proceedings of the Ninth Interface Symposium on Computer Science and Statistics* (D. C. Hoaglin and R. E. Welsch, Eds.), Prindle, Weber and Schmidt, Boston, 1976, pp. 44-51.
- Tarter, M. E. A description of the GRAFSTAT clinical life table options. In: *Proceedings of the Statistical Computing Section, The American Statistical Association*, Washington, DC, 1978, pp. 44-58.
- Tarter, M. E. Modular implementation of graphical curve estimation procedures. In: *Proceedings of the Statistical Computing Section, The American Statistical Association*, Washington, DC, 1981, pp. 69-78.
- Tapia, R. A., and Thompson, J. R. *Nonparametric Probability Density Estimation*. Johns Hopkins University Press, Baltimore, 1978, pp. 21-22.
- Quesenberry, C. P., and Starbuck, R. R. On optimal tests for separate hypotheses and conditional probability integral transformations. *Commun. Statist. Theory Methods* A5: 507-524 (1976).
- Tarter, M. E., and Kowalski, C. J. A new test for and class of transformations to normality. *Technometrics* 14: 735-744 (1972).
- Kendall, M. G. *Rank Correlation Methods*. Hafner, New York, 3rd ed., 1962.
- Tarter, M. E., and Kronmal, R. A. An introduction to the implementation and theory of nonparametric density estimation. *Am. Statistician* 30: 105-112 (1976).
- Tarter, M. E. Comment on Good and Gaskins. *J. Am. Statist. Assoc.* 75: 63-66 (1980).