STATISTICS IN MEDICINE
Statist. Med. **18**, 975–987 (1999)

A VISUALIZATION OF CROSS-OVER DATA USING LINEAR FUNCTIONS

WILLIAM E. MILLER*

National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 1095 Willowdale Rd., Morgantown, WV 26505-2888, U.S.A.

SUMMARY

Previous work has shown that cross-over trial data can be analysed using within-subject linear functions. Scores that result from linear functions are graphed in quantile comparison plots in order to visualize the differences between factor levels. An example suggests how this visualization can be used to identify outliers or to provide a more specific interpretation of results. Additional examples indicate how this approach can be used to track carry-over differences or interaction effects. This article is a US Government work and is in the public domain in the United States.

1. INTRODUCTION

Cross-over designs have a wide use in medical research, including pharmacological and pulmonary function studies. More complex designs have been introduced which allow for the measurement of the direct-treatment effects without assuming the absence of carry-over effects, but they require some strong assumptions about the carry-over effects.[1,2] Even if the assumptions are met, a traditional analysis still carries the same burden of assumptions as other designs with three or more repeated measures. Alternative approaches, which require fewer assumptions, have been presented by Koch[3] for a two-period cross-over design and by Hafner *et al.*[4] for a three-period cross-over design. Both papers recommend the use of within-subject linear functions when there are violations of the assumptions of normality or of sphericity. This method is akin to the examination of factors in a multivariate analysis through the use of linear combinations of the dependent variables.

Examples are presented which demonstrate how linear functions are applied in cross-over trials with two sequences. However, the motivation for using these linear functions will be to use the resulting scores in standard and detrended quantile comparison plots. By using the linear functions in conjunction with these plots, insight into factor differences is provided which is not easily apprehended by traditional methods. Moreover, this visualization can lead to a more specific interpretation of results, detect outliers, or provide a check on model assumptions.

* Correspondence to: William E. Miller, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, 1095 Willowdale Rd. Morgantown, WV 26505-2888, U.S.A. E-mail: wem0@cdc.gov

Contract/grant sponsor: National Institute for Occupational Safety and Health, U.S.A.

Received November 1997
Accepted June 1998

Table I. The source table with partial sums-of-squares for the analysis of the ABB and BAA sequence groups in the systolic blood pressure study. The treatment difference is statistically significant and a least-squares estimate of the difference indicates that the blood pressure is about 6 mm Hg lower for treatment B than for treatment A. Estimates of this treatment difference are also found using three linear functions

Source table and estimates of treatment difference

Source	D.F.	Sums of squares	Mean squares
Subject	48	31022·9	646·3
Period	2	395·1	197·6
Treatment	1	1133·6	1133·6
Carry-over	1	173·1	173·1
Error	94	16192·1	172·3

Least-squares estimate of treatment difference (A − B) 5·92 (standard error = 2·31)

Linear function estimates of treatment difference ($A - B$)

Linear function	Associated estimate
L_1: 1/2 (I + III) − II*	1·91 (SE = 4·18)
L_2: 1/2(I + II) − III*	9·93 (SE = 4·93)
L_3: 1/2I − 1/4(II + III)	5·92 (SE = 2·35)

* Assumes no difference between carry-over effects for the two sequences

2. EXAMPLE

A data set is found in Jones and Kenward,[5] which come from a study by Ebbutt.[6] It concerns the treatment of systolic blood pressure in hypertensive subjects. This trial has a three-period cross-over design with two treatments, labelled A and B, with four sequence groups, which are labelled ABB, BAA, ABA and BAB. For instance, subjects in sequence ABB received treatment A during the first six-week period, followed by treatment B in the second period, and treatment B again in the third period. All of the subjects had a baseline measurement and measurements for the three six-week treatments, but there were no washout periods between the treatments for ethical reasons.

Following the example of Jones and Kenward, we first use only the data from sequences ABB and BAA for illustration, ignoring the baseline measurements. The source table for the within-subjects analysis of variance is contained in Table I. Below the source table is the least-squares estimate of 5·92 millimetres for the treatment difference (A − B), which has an associated probability of 0·01. In other words, the blood pressure of a subject was, on average, approximately 6 mm Hg lower when treatment B was used.

Hafner *et al.* show how linear combinations of the first-period (I), second-period (II), and third-period (III) responses can be used to estimate the treatment differences and other factor

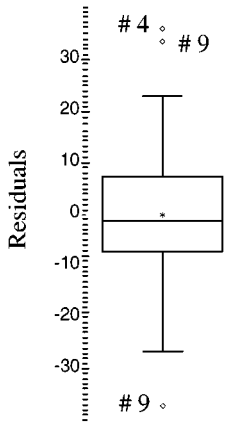

Figure 1. Box plot of the residuals from the analysis of variance of the ABB and BAA sequence groups in the systolic blood pressure study. The labels identify two subjects with responses which had relatively poor fits with respect to the model: subject 4 of the ABB sequence group and subject 9 of the BAA sequence group

differences, as long as we assume (as we do here and elsewhere) that carry-over effects are 'simple' (last for at most one period) and also that the carry-over effect of a treatment into itself is equivalent to other carry-over effects. One example of such a linear function is $L_1 = 1/2(\mathrm{I} + \mathrm{III}) - \mathrm{II}$. By applying it to the triad of responses from each subject, one linear function score is obtained for each of the subjects. The difference of the means of these scores for the two sequence groups is 1·91, which represents an estimate of the treatment difference. It can be shown, by examining expected values, that there are other linear functions which estimate the same treatment difference. Two other linear functions which estimate this difference are $L_2 = 1/2(\mathrm{I} + \mathrm{II}) - \mathrm{III}$ and $L_3 = 1/2\mathrm{I} - 1/4(\mathrm{II} + \mathrm{III})$. The results of applying these three linear functions are also shown in Table I.

The third linear function, L_3, produces an estimate equivalent to the least-squares estimate. It also has the smallest variance estimate of the three linear function estimates. Hafner *et al.* state that the linear function with the smallest variance is equivalent to the least-squares estimator. In addition, unlike the other linear functions it is not necessary to assume the absence of (simple) carry-over effects to order to use L_3 to estimate the treatment difference. The remainder of the paper focuses on the 'least-squares' linear functions.

The distribution of the residuals from the analysis of variance is shown in the box plot of Figure 1. The box plot identifies three large residuals, one resulting from a response from subject 4 of the ABB sequence group and the other two resulting from subject 9 of the BAA sequence group. The large residuals for subject 9 result from the first response (using treatment B) being smaller than that predicted by the model and the third response (using treatment A) being larger than that predicted by the model.

3. PLOTTING LINEAR FUNCTIONS

The quantile comparison plot, or quantile-quantile (q-q) plot, was introduced by Wilk and Gnanadesikan,[7] and, as the name implies, it involves plotting the quantiles of one (continuous)

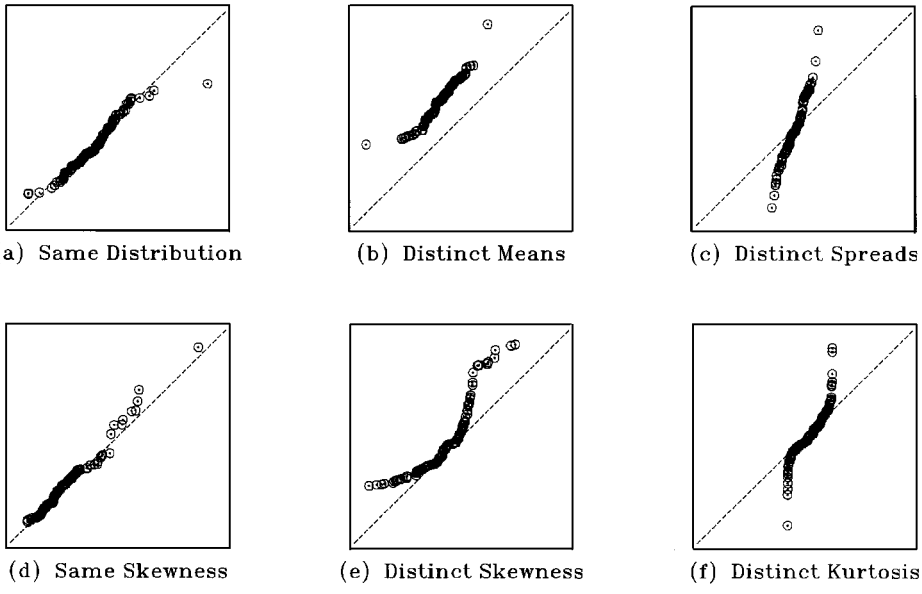

Figure 2. Six quantile comparison plots are displayed for pairs of computer-generated data sets with 100 observations each. In general, two data sets from distributions having similar shapes will produce a quantile comparison plot which approximates a straight line. If the distributions also have similar spreads, then the line will be approximately parallel to the reference line. Comparisons are made for: (*a*) two normally-distributed data sets which do not differ (except for random variation); (*b*) two normally-distributed data sets which differ with respect to their means; (*c*) two normally-distributed data sets which differ with respect to their variances; (*d*) two skewed data sets which do not differ (except for random variation); (*e*) a positively-skewed data set versus a normally-distributed one (reflecting this plot about the reference line would result in a graph depicting the first distribution as negatively-skewed relative to the second distribution); (*f*) a data set whose distribution is leptokurtic (having lighter tails) compared to a normally distributed one (reflecting this plot about the reference line would result in a graph depicting the first distribution as being heavy-tailed, or platykurtic, relative to the second distribution)

distribution against the quantiles of another. It can be used (i) to compare two theoretical distributions, (ii) to compare the quantiles of a sample (an empirical distribution) to those of a theoretical distribution, or (iii) to compare the quantiles of two samples. If two samples are of the same size, the graph is produced by plotting the order statistics of one sample against those of the other. When the samples are not of the same size, then some sample quantiles must be estimated using a method of interpolation.

Figures 2(*a*)–(*f*) provide a brief illustrative key for some patterns that emerge in quantile comparison plots using computer simulations. As these figures indicate, two samples from distributions having similar shapes will generally result in a plot of data points which approximates a straight line. If the distributions have similar spreads, then that line will be approximately parallel to the diagonal reference line. If the central tendencies are also similar, then the plotted points will roughly coincide with the reference line. Figures 2(*e*) and 2(*f*) show examples of patterns that result when two distributions differ with respect to shape.

It is evident from Figures 2(*a*)–(*f*) that the quantile comparison plot has a heuristic appeal for the data analyst. If two distributions have approximately the same shape and spread, then this will generally lead to plotted points being approximately the same vertical distance to the

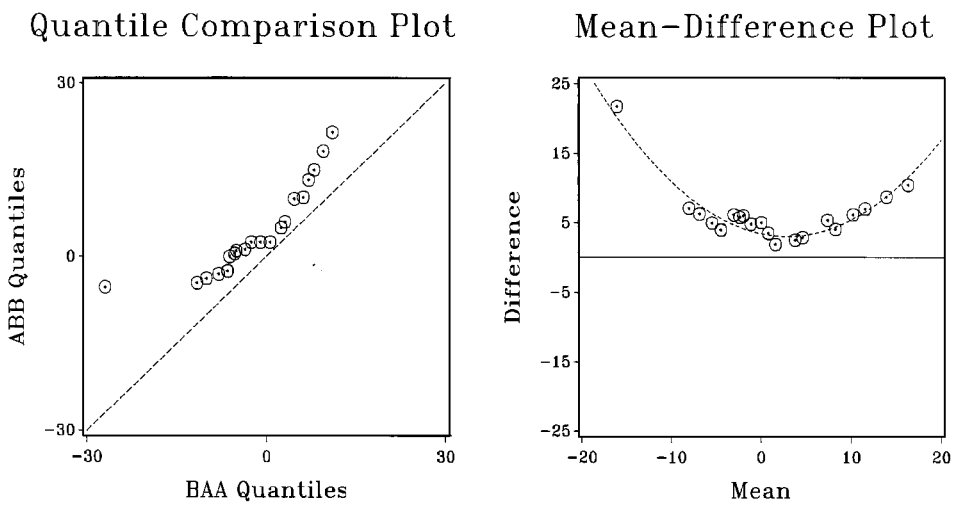

Figure 3. The quantile comparison and mean–difference plots to visualize the treatment differences by using the L_3 linear function scores of sequences ABB and BAA in the systolic blood pressure study. The mean of the y co-ordinates in the mean–difference plot is approximately equal to the least-squares estimate of the treatment difference (A − B). The responses of subject 9 of the BAA sequence group account for the large difference between the lowest sample quantiles. Subject 9 is identified in the box plot of Figure 1 as having two large residuals resulting from the analysis of variance

reference line. This means that the difference between respective quantiles will be fairly uniform across the distributions of the samples. Therefore, the plot indicates the appropriateness of representing the differences between two samples by a single statistic, such as the difference between means. However, substantial departures from the reference line can occur by chance, particularly when the sample sizes are small.[8] There can also be erratic patterns in the tails, even when sample sizes are adequate.

For the Ebbutt data there are 22 subjects in the ABB sequence group and 27 subjects in the BAA sequence group. Using the linear function L_3, scores are calculated from the three responses of each of the 22 subjects in the first sequence group. These 22 ordered scores are then plotted against the 22 interpolated quantiles derived from the 27 scores of the second sequence group, where the method of interpolation is the one suggested by Hyndman and Fan.[9] This plot is shown in Figure 3. However, because the vertical distances to the reference line are more difficult to compare in the quantile comparison plot, we also produce a detrended version of the quantile comparison plot, also known as Tukey's mean–difference plot.[10] This is a plot of the differences between the paired quantiles against their means. It is comparable to rotating the quantile comparison plot clockwise through 45° and then expanding the vertical axis. In the mean–difference plot of Figure 3 the overall mean of the ordinates (y co-ordinates) for all of the plotted points is approximately 5·93, nearly equal to the least-squares estimate of the treatment difference. When interpolation is not necessary, there is equality.

The mean–difference plot of Figure 3 shows that the difference between the lowest sample quantiles is about 20 millimeters of mercury, much larger than the differences between other paired quantiles. The univariate distributions for the two sample quantiles can be pictured by mentally projecting the plotted points against the two axes of the quantile comparison plot of

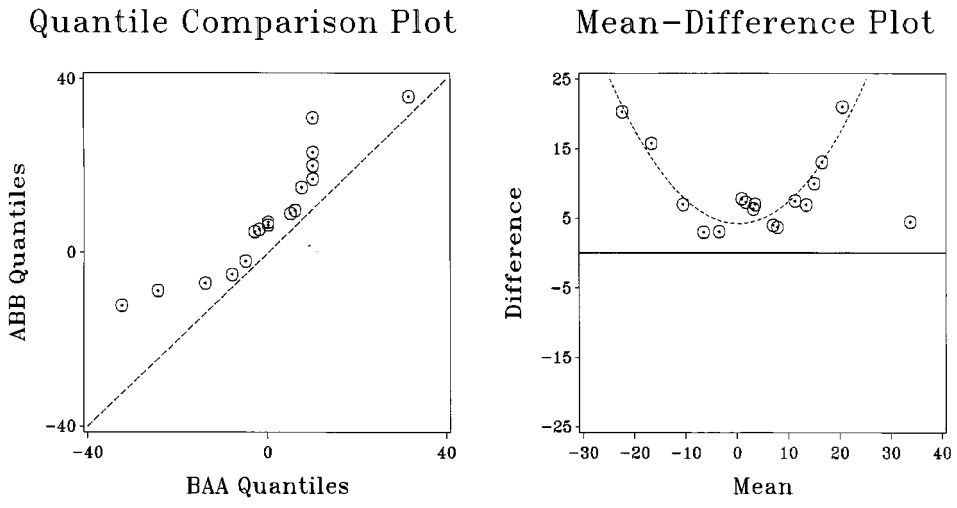

Figure 4. The quantile comparison and mean–difference plots to visualize the treatment differences by using the L_4 linear function scores of sequences ABA and BAB in the systolic blood pressure study. The mean of the y co-ordinates in the mean–difference plot is approximately equal to 8·7, the least-sqaures estimate of the treatment difference (A − B)

Figure 3. This shows that the lowest sample quantile of sequence BAA is an extreme value. Further investigation (not shown here) shows that this extremity is caused by the previously identified results for subject 9 of sequence BAA.

A comparison with the key of Figure 2 would suggest differences in skewness, but another interpretation is more useful to researchers. In the mean–difference plot the differences between the sample quantiles are smaller near the centres of the distributions. This suggests that subjects in the study with average blood pressure tended to exhibit smaller differences in their responses to the two treatments than subjects with relatively low or high blood pressures. The quadratic regression line provides an adequate fit to the plotted points of the mean–difference plot of Figure 3.

It is possible to validate this pattern by using the results for the other two sequence groups in the study. The least-squares linear function for sequences ABA and BAB is $L_4 = \mathrm{I} - 1/2(\mathrm{II} + \mathrm{III})$ and the resulting scores are plotted in Figure 4. The mean–difference plot of Figure 4 also shows a tendency for subjects with relatively low or high blood pressures to have larger differences in their responses to the two treatments, with the exception of the rightmost plotted point. The quadratic regression line has been fitted to all but this point, which lies far from the prediction line because of responses from subject 1 of sequence BAB. Two large residuals due to this subject were found from the analysis of variance (not shown here), and they resulted from the first response (using treatment B) being larger than that predicted by the model and the third response (also using treatment B) being smaller than that predicted by the model.

This approach can also be used for other effects in the model. For instance, using the results for sequences ABB and BAA, a quantile comparison plot (not shown here) was produced using the linear function scores for the carry-over effects. The one clear outlier in the plot was associated with the results of subject 4 of the ABB sequence group. This is the other subject who is identified in the box plot of Figure 1.

4. ADDITIONAL EXAMPLES

A method for finding a least-squares linear function is found in the literature for cell-means or full-rank linear models. The connection between the hypothesis matrix δ in the effects model hypothesis

$$\delta\beta = 0 \tag{1}$$

and the corresponding hypothesis matrix H in the cell-means model hypothesis

$$H\mu = 0 \tag{2}$$

is given by Miller,[11] using results from Speed,[12] as

$$H = \delta X^{-} \tag{3}$$

where X^{-} is the generalized inverse of the design matrix X of the effects model, and where δ is also the coefficient matrix for the estimable functions. The elements of H will contain the coefficients of the least-squares linear function.

4.1. Insulin Mixtures

The text of Jones and Kenward[5] contains an example of a two-treatment four-period cross-over study by Ciminera and Wolfe.[13] The purpose of the trial was to compare the blood sugar levels of two insulin mixtures among 22 rabbits. The two sequences of treatments A and B are ABAB and BABA. The rabbits were injected with one of the insulin mixtures each week for four weeks and repeated blood samples were taken at five time periods after the injection: immediately after the injection (time = 0), and 1·5, 3·0, 4·5 and 6·0 hours later.

Using the method for finding linear functions outlined above, the linear function for the treatment difference (A − B) is calculated as $L_5 = 1/4(4*\mathrm{I} - \mathrm{II} - 2*\mathrm{III} - \mathrm{IV})$ where I to IV represent the first- to fourth-period responses for a subject. The linear function for the carry-over difference for treatments A and B is $L_6 = \mathrm{I} - \mathrm{III}$. As an illustration, L_6 is applied to the responses immediately after the injection (time = 0), and separately to the responses at 1·5, 3·0 and 4·5 hours later. In Figure 5 the scores are graphed in quantile comparison plots for the four time periods and this allows us to track the residual effects of the treatments over time. A common scale of measurement is used and it shows a stretching of the data points along the reference line as time increases. This increase in the variability over time also extends into the 6-hour measurements, not shown here. However, the carry-over differences for the treatments appear to be small except for the 3·0-hour measurements. Its pattern suggests a possible difference in the shape or variation for the carry-over effects of the two treatments, although there are too few subjects to show a definite pattern.

4.2. Classifications of Chest X-rays

A more complicated example comes from a recent study[14] sponsored by the International Labour Office (ILO). The International Classification of Radiographs of Pneumoconioses[15] is a classification scheme adopted by the ILO which is composed of 12 ordered categories, from '0/−' '3/+'. The increasing categories of this 'profusion' classification correspond to the observation of more frequent shadows, or opacities, in chest X-rays. A certified physician classifies the profusion of a chest X-ray (or radiograph) into one of the categories by comparison with a set of 22 standard radiographs.

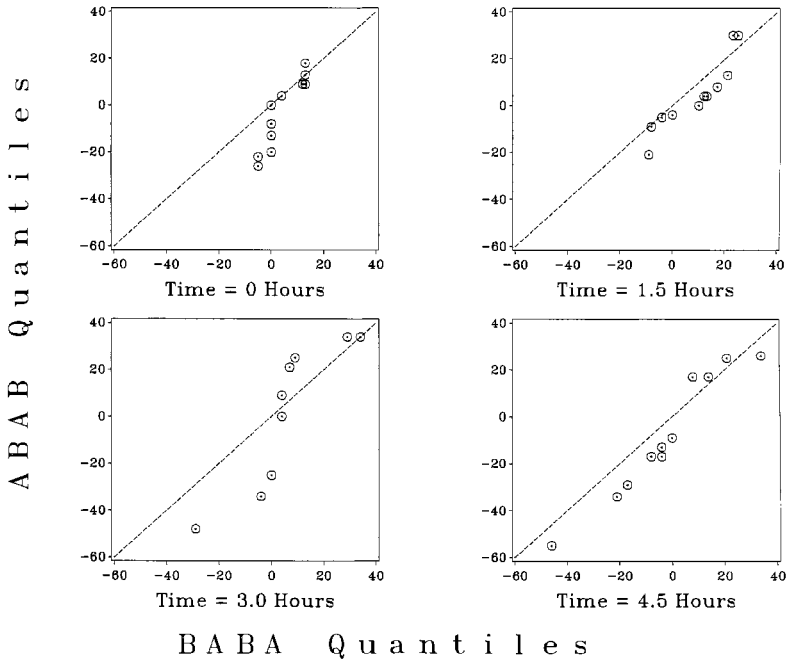

Figure 5. Quantile comparison plots to visualize the difference in the carry-over effects for the two treatments A and B of the blood sugar study by Ciminera and Wolfe. Each plot depicts the differences between the carry-over effects during one of four time periods after the injection

The ILO has been considering a modification of the 22 standard radiographs which would make them easier to use, and the study was conducted in order to assess how these changes might alter the use of the classification scheme. During the first two periods of the study, physicians classified a set of 120 X-rays twice, once using the current full-size ILO (1980) set of standards and once using a set of standards which showed sections of the full-size standards within four quadrants of a radiograph. This modified or 'Quadrant' set of standards comprised 14 radiographs. The first two periods of the trial define two sequence groups of physicians, IQ and QI, depending on whether a physician used the ILO standards or the Quadrant standards in the first period. In the third period of the trial each physician classified the profusion of 60 of the 120 X-rays using one set of standards and then classified the other 60 X-rays using the other set of standards. The third period further defines two subsequences of classifications for each sequence of physicians: (1) subsequences IQI and IQQ for the sequence IQ physicians, and (2) subsequences QIQ and QII for the sequence QI physicians.

The opacities found on chest X-rays are known to have a predominantly rounded shape in studies involving exposure to coal dust and a predominantly irregular shape in studies involving asbestos exposure. The physicians also classified the opacities of the 120 X-rays into these shape categories, and the initial analysis suggested that their profusion classifications depended, to some extent, on whether they specified the opacities as having a rounded or an irregular shape. For the present analysis, the triad of classifications of an X-ray by a physician is categorized as 'rounded'

if at least two of the three classifications specified the shape as rounded. Similarly, a triad of readings is categorized as 'irregular' if at least two of the three classifications specified the shape as irregular.

In addition, a peculiar result from the initial analysis was that profusion classifications in the IQQ subsequence tended to be much higher in the third period than in the second period, even though the physicians used the same set of standards on both occasions. This standards-by-period interaction might be interpreted as a learning effect that occurred when physicians classified the films a second time using the modified standards. In the language of cross-over designs, this learning effect can be viewed as the carry-over effect of a treatment effect into itself. In order to investigate this result, we restrict our attention to a comparison of subsequence IQQ to subsequence QII, its 'dual sequence'.[5] These subsequences include 19 and 18 physicians, respectively. The lowest two categories, '0/ − ' and '0/0', have been combined into one category and integer scores of 0 to 10 assigned to the eleven ordered categories. The means of these integer scores are displayed in Table II for each reader, period and shape combination.

The differences between the classification levels for the standards (Quadrant − ILO) is examined by applying the linear function $L_7 = 1/2(\mathrm{II} - \mathrm{I})$ to the results in Table II for the first and second periods of the trial. The linear function scores, stratified by shape, are plotted in the first row of Figure 6. The mean–difference plots suggest that, during the first two periods of the trial, the classifications tended to be slightly higher using the Quadrant standards when the shape of the opacities was identified as irregular. On the other hand, the classifications tended to be slightly higher using the ILO standards when the shape was specified as rounded. The means of the ordinates for the two plots are approximately 0·14 and −0·19. Therefore, the net effect amounts to about 14 higher categories per 100 X-rays when using the Quadrant standards to classify X-rays with predominantly irregular shapes, and about 19 higher categories per 100 X-rays when using the ILO standards to classify X-rays with predominantly rounded shapes.

The difference of the carry-over of the Quadrant standards into itself and the carry-over of the ILO standards into itself can be estimated by applying the linear function $L_8 = \mathrm{III} - \mathrm{II}$. The corresponding mean–difference plots, stratified by shape, are in the second row of Figure 6. The means of the ordinates for these plots are 0·25 and 0·45. Therefore the average change from the second classification to the third classification of an X-ray is estimated as about 25 to 45 higher categories per 100 X-rays for the Quadrant standards than for the ILO standards. The change in classification that occurs later in the trial is a relatively larger effect than the difference between the classification levels of the two standards, as judged by the first two periods of the trial. Also, the plotted points in the second row have a larger spread, which suggests that the variation associated with the change in classification is also larger.

The patterns in Figure 6 are similar in the two plots for linear function L_8, except for a slightly larger shift towards higher classifications of rounded shapes using the Quadrant standards. If we accept these changes as indicating a learning effect, then a comparison of the plots based on the linear function L_8 to those based on L_7 could suggest that differences between the standards begin to emerge at an earlier period of classification involving irregularly-shaped opacities. The larger adjustment that occurs later for the classifications involving rounded shapes would then indicate that, in the long run, there may be higher classifications using the Quadrant standards, regardless of the shape classification. However, it appears that physicians with average classification levels can be excluded from these conclusions, since their classifications show relatively small differences between the standards and small adjustments over time.

Table II. Means for the chest X-ray classifications for 37 physicians, stratified by shape and period. These means are used to calculate the linear function scores which are plotted in Figure 6. Note that means for different readers are not necessarily based on the same set of X-rays, but the linear functions L_7 and L_8 are contrasts which remove the effects due to the different X-rays. However, the linear function for the carry-over effects (I + II) in the first and second periods would not remove the effects due to the different X-rays. For similar reasons, inferences based on marginal means would be misleading

Sequence	Physician	Irregular shapes Period			Rounded shapes Period		
		1	2	3	1	2	3
IQQ	1	3·94	2·68	2·82	5·40	4·53	4·87
	2	3·17	2·97	3·60	5·54	5·00	5·08
	3	3·32	3·00	4·14	5·08	4·17	5·67
	4	2·29	2·62	2·48	3·95	3·89	3·95
	5	3·15	3·60	3·25	3·35	3·41	3·56
	6	2·58	2·70	2·64	4·82	4·18	5·00
	7	3·09	2·64	3·36	3·13	3·08	3·50
	8	2·91	4·00	4·09	3·45	3·86	4·10
	9	2·92	2·81	3·46	4·32	4·37	4·32
	10	2·70	3·55	3·15	4·07	5·07	4·73
	11	3·38	4·00	3·95	3·88	3·79	4·21
	12	2·53	2·43	2·30	5·00	3·58	4·25
	13	1·28	2·17	3·11	2·50	3·33	3·42
	14	2·57	2·91	3·66	3·89	3·72	5·06
	15	3·53	3·42	4·84	4·30	2·83	4·00
	16	3·26	2·32	3·42	3·75	3·33	3·83
	17	2·33	3·58	2·92	5·14	4·29	5·14
	18	2·43	2·79	1·82	3·20	3·80	2·53
	19	4·11	3·22	4·22	4·14	3·70	3·86
QII	1	3·43	4·14	3·21	4·50	4·94	4·44
	2	3·46	4·23	4·23	3·87	4·13	4·16
	3	3·08	3·08	3·15	4·69	5·56	5·25
	4	2·67	2·33	1·80	4·54	4·08	4·46
	5	3·10	2·73	2·63	5·14	5·79	4·71
	6	2·32	2·00	1·94	4·83	4·92	4·58
	7	2·84	2·12	2·48	4·96	4·56	4·32
	8	2·87	2·91	2·65	4·89	4·58	4·84
	9	3·44	2·97	3·28	5·90	5·70	6·10
	10	3·50	2·85	3·22	6·22	6·22	6·44
	11	4·06	3·34	3·56	5·75	5·25	5·42
	12	2·59	2·86	3·00	4·23	5·00	5·15
	13	3·10	2·48	3·14	3·70	3·65	3·65
	14	2·83	3·13	3·29	3·65	4·20	4·45
	15	3·32	2·53	2·74	4·36	3·09	3·45
	16	3·50	3·46	2·17	3·82	3·95	3·00
	17	3·26	2·37	3·68	4·00	4·31	4·92
	18	3·46	3·92	4·31	4·15	5·41	4·63

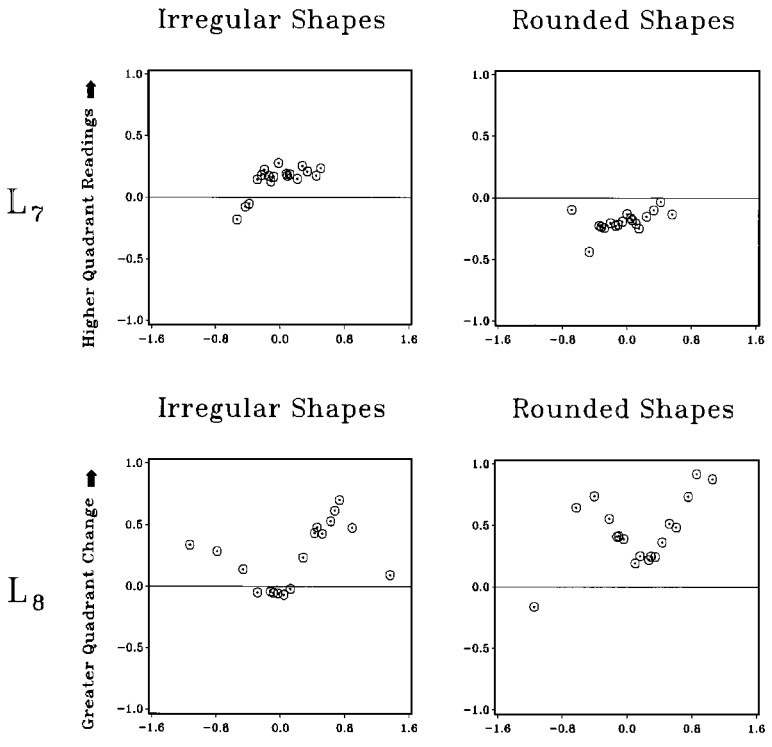

Figure 6. Mean–difference plots for the two linear functions L_7 and L_8 which are used in the analysis of the chest X-ray study. The plots based on L_7 provide a visualization of the differences in classification levels using the two sets of standards during the first and second periods of the trial. The plots based on L_8 provide a visualization of the differences between the standards with respect to the adjustments that the physicians made in their classifications from the second to the third periods of the trial. The results are stratified by whether the profusion classifications are associated with rounded or irregular shape classifications

5. DISCUSSION

Three examples were given which used linear function scores in conjunction with quantile comparison plots. In the blood pressure example, a box plot identified three large residuals from the analysis of variance, but quantile comparison plots also indicated how the outliers would affect estimates of factor differences. Quantile comparison plots also showed differences in shape between the distributions of the linear scores which suggested that treatment differences depended on location within the distribution.

The blood sugar example showed how these plots might be used to track the carry-over effect over time. When there is a sufficient number of subjects, it might be possible to track the carry-over effect in order to determine an optimal time period in which to compare the treatment effects.

The chest X-rays example allowed us to compare the estimated standards effects that occurred earlier in the trial to an interaction which occurred later in the trial. Like the first example, it calls

attention to a situation where an 'average' difference may not be typical of the distribution of differences.

Some of the results using these methods could have implications for model misspecification and the use of these methods will need to be accompanied by some attention to the model assumptions (see the texts by Jones and Kenward[5] and by Senn.[16]) Also, as the second example showed, these methods will have limited applicability in trials with few subjects. It would be desirable to have at least 30 subjects in each sequence group to determine questions of shape, although the examples presented here show that it is possible to validate some broad patterns in studies with fewer subjects.

It is known that, in order to perform a fixed effects analysis of variance with one response, it is not necessary to access the data. All that is required are the means, the standard deviations and the sample sizes for each factor combination of the model. If there are multiple responses, then the covariance matrix is also needed. This suggests the tendency of modelling to shift attention away from the data towards the estimation of parameters. This can sometimes leave a vague connection between the descriptive results and the modelling results in the minds of researchers who work with statisticians. The methods presented can also, in some cases, illuminate the results of an analysis of variance and may make the presentation more cogent.

An SAS™ macro, which constructs quantile comparison and mean–difference plots for two samples, is available from the author. An example of SAS IML®[17] code, which is used to calculate a least-squares linear function, is also available.

ACKNOWLEDGEMENTS

The author thanks Kathleen Fedan and Martin Petersen of NIOSH, and the referees for their helpful comments. This research was supported by the Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Morgantown, West Virginia, U.S.A.

REFERENCES

1. Senn, S. J. 'Is the "simple carry-over" model useful', *Statistics in Medicine*, **11**, 715–726 (1992).
2. Laska, E., Meisner, M. and Kushner, H. B. 'Optimal crossover designs in the presence of carryover effects', *Biometrics*, **39**, 1087–1091 (1983).
3. Koch, G. G. 'The use of non-parametric methods in the statistical analysis of two-period change-over design', *Biometrics*, **28**, 577–584 (1972).
4. Hafner, K. B., Koch, G. G. and Canada, A. T. 'Some analysis strategies for three-period change-over designs with two treatments', *Statistics in Medicine*, **7**, 471–481 (1988).
5. Jones, B. and Kenward, M. G. *Design and Analysis of Cross-over Trials*, Chapman and Hall, London, 1989.
6. Ebbutt, A. F. 'Three-period crossover designs for two treatments', *Biometrics*, **40**, 219–224 (1984).
7. Wilk, M. B. and Gnanadesikan, R. 'Probability plotting methods for the analysis of data'. *Biometrika*, **55**, 1–17 (1968).
8. Hoaglin, D. C. 'Using quantiles to study shape', *in* Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (eds), *Exploring Data Tables, Trends, and Shapes*, Wiley, New York, 1985, pp. 417–460.
9. Hyndman, R. J. and Fan, Y. 'Sample quantiles in statistical packages', *American Statistician*, **50**, 361–365 (1996).
10. Cleveland, W. S. *The Elements of Graphing Data*, Revised edn, Hobart Press, Summit, New Jersey, 1994.
11. Miller, R. L. 'The cell means model as an analytical tool for evaluating SAS GLM Type III and IV sums of squares for linear models with missing cells', in *Proceedings of the Sixth Annual SAS Users Group International Conference*, V. 6, 1981, pp. 547–554.

12. Speed, F. M. 'A new approach to the analysis of linear models', Ph.D. Dissertation, Texas A.M. University, 1969.
13. Ciminera, J. L. and Wolfe, R. K. 'An example of the use of extended cross-over designs in the comparison of NPH insulin mixtures', *Biometrics*, **9**, 431–446 (1953).
14. NIOSH, 'A trial of additional composite standard radiographs for use with the ILO International Classification of Radiographs of Pneumoconioses', Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication HETA 93-0340, 1997.
15. *Guidelines for the Use of ILO International Classification of Radiographs of Pneumoconioses*, Revised edn 1980, Occupational Safety and Health Series, No. 22 (revised), International Labour Office, Geneva, 1980.
16. Senn, S. J. *Cross-over Trials in Clinical Research*, Wiley, Chichester, 1993.
17. SAS Institute. *SAS IML® User's Guide*, *Release* 6.03 *Edition*, SAS Institute, Cary, N.C., 1988.