A STATISTICAL STUDY OF VARIATION IN SURGICAL DRESSINGS

BY R. MAXWELL SAVAGE

From the Laboratory of S. Maw Son, Sons Ltd.

Received July 12, 1954

IN a recent book on industrial statistics[1] a chapter opens with the sentence "The subject of specifications is involved, and for the majority of industrial products, considerably more evidence and data are required before rational specifications based on statistical principles can be drawn up." This statement certainly applies to surgical dressings, and the object of this paper is to examine the specifications now in use using statistical principles and to compare them with the articles produced by their guidance, and finally to make suggestions for improvements.

For this purpose, it has been necessary to adopt a critical attitude as a part of the technique. It is therefore important at the outset to emphasise that, in the main, surgical dressing specifications are, like other official standards, no more than codifications of existing satisfactory practice. There is in general no demand for the raising of the standards and if an article frequently does not appear to comply with a specification it will be shown that it is at least as likely that the specification has fallen short of the perfect reflection of the normal facts it should ideally be, as that the dressing itself has failed to reach a clearly defined objective standard. The paper is based on the hypothesis that the existing articles are on the whole correct, whether this is true or not.

The necessity of reaching agreement on the properties of variable material presented itself to the textile industry long before modern statistical methods existed. The solution reached was the gradual evolution of traditional interpretations of the various terms, with the official testing houses standing as referees in cases of dispute. The fact that this system works does not make it less important to examine modern methods for specifying variable materials, for those unversed in the tradition are apt to expect the written specifications to do more in the way of definition than at present they can possibly do. It may be that newer methods have something to contribute, for it will be shown in this paper what different interpretations can be placed upon apparently precise descriptions of the simplest properties. Published information on the statistical variation in surgical dressings appears to be lacking. In the British Pharmaceutical Codex the fact of variation is only recognised by the existence in some instances of limits, and these are insufficient for the present purposes. It therefore seems essential in order to criticise the specifications and the customs of the trade, to report the results of tests which show what variations do occur.

Scope. The only cases considered are purely textile specifications for some cloths of the B.P.C., that is to say, threads per inch, and weights per unit area. Extensions to other cases are mere matters of application of the same principles.

Methods. In some cases, samples of dressings were bought in the open market so as to provide information about the prevailing properties of dressings. Detailed investigations were carried out on repeated deliveries of bulk material over a long period, many thousands of yards being sampled, or for special purposes, by exhaustive examination of particular portions of cloth. Conditioned weights were obtained by drying the samples completely and adding 8·5 per cent. to the dry weight. Not all the work was done on conditioned material, for in some comparisons only relative results matter. Such cases are indicated in the text. The weights of cloth are expressed as g./100 sq. cm., a 10 cm. × 10 cm. template being used for cutting the specimens. Weights per unit area determined by measuring and weighing pieces of different sizes are given in the same units, but it is invariably stated when this method was adopted.

Threads per inch.

The B.P.C., Appendix X, states that "In fabrics of open texture, the number of threads in a length of 10 in. should be counted. . . ." In the various monographs, the number of threads is invariably stated as "Average not less than——." At first sight, these statements together appear to be reasonably precise.

The British Standards Handbook No. 11, 1949 Edition[2], described 4 methods for determining the threads per inch in cloth. In the preface, it is stated that the testing methods" . . . represent current technique." This is for the present case more of an aspiration than a fact, for during the preparation of this paper, two experienced weaving executives have independently failed at first to understand how fractional numbers of threads per inch can be obtained, in spite of the fact that all the 4 methods and the B.P.C. Statement necessarily lead to such results. There was not unanimity among other informed people in the dressings trade. Views were obtained ranging from those which regarded one thread short in 10 in. as a deficiency, to one which was to the effect that any number over the next whole number lower than the specified limit was correct, provided that no place in the cloth contained less than this. It is obvious that 3 kinds of error (definition, determination and interpretation) are involved in the confusion, and an attempt to clarify the differences will now be made for they are unlikely to be resolved before they are clear.

The four methods of the Handbook are—

(1) Traversing thread counter. In this method the number of threads in a length of 5 to 10 inches are counted under a lens.

(2) Dissection of a measured length of cloth and counting the separated threads.

(3) The use of a diffraction grating having a certified number of lines, with correction for the number of interference bands produced.

(4) The 1-in. counting glass, in which the number of spaces, visible in the glass, with an estimate of a fractional space, is repeatedly counted and an average taken.

Besides these four 'official' methods, there is another.

(5) The traditional use of the 1-in. counting glass. The glass is placed anywhere on the cloth and the visible threads are counted. This can, in different hands and for different purposes, give results which are near either the maximum or the minimum possible, and provides the widest possible field for disagreement. It is still in general use, and must be considered here.

Since no reference is made in the B.P.C. to the method of determination 4 of them have been examined, but before presenting the results, it is essential to discuss the actual wording of the B.P.C.

"*Open Texture.*" The B.P.C. requires that in cloths of open texture, the average number of threads in 10 in. shall be counted. Open texture is not defined. The Handbook states that the counting glass method is unsuitable for cloths with less than 25 threads per inch so perhaps this can be taken as a limit. It is clearly by no means a matter of indifference as to whether 1 in. or 10 in. of cloth is taken for the determination. In the experiments below the count was 27, but this was considered so near to the limit that it should be regarded as an open cloth.

"*Average.*" The following alternative interpretations are possible.

(1) The average may be at one extreme, that of the whole of the manufacturers' production, or, at the other extreme, that of any single piece in the hands of the analyst. A statistical relationship always exists between these numbers, but the basis for it is nowhere stated.

(2) The various methods of determination may yield different averages. It will be shown that the unofficial method does in fact do so.

(3) Even the methods in the Handbook are not carried out on the same lengths of cloth. Methods 1 and 2 use a single length of 1 in. repeated 5 or 10 times. Method 3, the average of at least 3 separate 8-in. lengths. Method 4, single 1-in. lengths repeated at least 10 times. Method 5, 1 in. length, without provision for repetition.

(4) No directions for random sampling occur. In Method 5, sampling is sometimes selective, according to the point of view of the operator, and the final result is an average of maxima or minima, which is not definitely excluded by the specification.

"*Not less than.*" This phrase has no particular meaning when applied to variable material in which every particle cannot be inspected, accepted or rejected, unless either (*a*) the limit is so far from the normal average that it merely excludes definite mistakes or fraud, (*b*) a statement of variation in statistical terms or their equivalent, accompanies the number. Neither of these conditions applies to the items considered in this paper. It must be emphasised here that on the hypothesis that dressings made, if this were possible, exactly to the B.P.C. specification are efficient, raising the average to bring them into class (*a*) does not indicate virtue in the manufacturer but is economic waste. The statistical implications are dealt with in a later section.

Methods of Determination.

In order to compare the various methods, a single piece of open-wove bandage cloth measuring about 40 in. square was taken, and 5 methods applied to it by 3 people. The doubts about the lengths and repetitions required were settled by the following arbitrary decisions.

Method 1. 10 separate 10-in. lengths of the cloth, chosen at random, were counted. The mean for each 10-in. length was considered as a single observation.

Method 2 was not used because it involves the destruction of the cloth.

Method 3. 3-in. gratings were used on 10 portions chosen at random, the average number of threads in each 3-in. length being considered as a single observation.

Method 4. 10 1-in. portions were chosen at random, each reading, with the estimated fraction, being considered as a single observation.

Method 5 (*a*) The glass was placed so as to enclose the maximum number of threads and every thread or part of a thread was counted. The glass was not placed at random, but efforts were made to choose the densest places, though in fact the cloth was sufficiently dense even to make this attempt of doubtful effect. (*b*) A similar procedure, but efforts were made to get the lowest result. Each reading, necessarily a whole number, was considered as a single observation.

The results were subjected to the analysis of variance which is given in the Appendix, and only the broad conclusions are stated here.

The mean was estimated most accurately by the 10-in. counts (Method 1) followed by the grating method (Method 3). The B.S. counting glass (Method 4) method came next. There is probably nothing to choose between these methods for accuracy. The difficulties of estimating the significance of variances based on small numbers make it simplest to consider all these official methods as equivalent. There is no question that the grating method is the quickest, thus confirming the statement in Shirley Test Leaflet No. A.C.5.[3] The traditional use of the glass to give a maximum, gave the only result with a mean which differed significantly from the others. This method also had the highest variance, but the real difficulty does not lie in the accuracy with which the mean is estimated, but in the fact that the traditional method gives a mean larger by 1·1 threads per inch than the others. This method could be used if everyone agreed upon it.

The size of the variance is of interest. In the case of the grating, the variance of means of 10 readings was 0·07 giving a standard deviation of 0·27. This means that about 1 per cent. of the results on a cloth woven to a manufacturer's bulk average of exactly 27·0 would be as low as 26·19 and less than 1 in 1000 results would be below 26·0. If this mathematical result is compared with one of the verbal opinions given above, it will be seen that it is practically identical—results between 26 and 27 will be common, but those below 26 will hardly ever occur. The other view, that even one thread short in 10 in., e.g., a count of 26·9 indicates a deficiency, may arise if the method of counting was the traditional maximum count by the glass, for a true count of 27 becomes over 28 by the traditional method and results below 27 would be very rare. It is also possible that there is a genuine misapprehension about the accuracy of cloth woven to any specification, for it is the experience of statisticians that manufacturers of variable material tend, in the absence of precise measures of variation, to over-estimate their own accuracy.

The personal factor is not important, for none of the 3 people produced significantly different results. This may be unexpected for there is an impression that the personal factor is large. It probably arises from the use of different methods by different people. There is evidence that the different methods did not "suit" the 3 persons equally (the interaction "persons/methods" was significant), but the practical importance of this is not great.

Wider results. Having investigated the counting methods on a single specimen of cloth, the next point is to examine cloth variation over a wider field. Table IV shows the 10-in. counts from 23 consecutive deliveries of cloth from the same source. The results can be taken as random samples of bulk production of many tens of thousands of yards, woven under conditions as identical as production methods allow. The mean was 27·49 and the standard deviation 0·48. The variance is larger than that from a single piece of cloth, because additional sources of variation, such as the variance between and within rolls, must be added to the variance from the results on a single small piece of cloth. If the distribution had been normal, there should have been about 4 values below 27, but in fact none were found. This cloth clearly passes the B.P.C. specification, partly because the mean is 0·49 above the limit, and partly because the distribution may possibly be slightly skewed in the direction of high values.

A still wider view is obtained from column 7 of Table VI in which the results are shown from 19 samples of various makes bought at random in retail shops. The mean is 27·00 and the standard deviation 0·97. From this it is reasonable to draw the conclusion that the trade target is 27, but examination of the individual readings shows two of 25·7 and 25·3 which, unless drawn from production with a very high standard deviation, are unlikely to come from cloth woven to a 27 average. Even the use of the traditional counting method could hardly explain this figure. On the whole, therefore, the trade appears to aim just above 27. 8 out of the 19 samples fall below the limit, which is about what one would expect from variable material made to a correct average.

It now becomes clear that the real standard, if our hypothesis is correct, is that the *bulk average* is not less than 27. This being so, no person holding a fragment of B.P.C. cloth has any reason to be sure that it will reach 27. If he has a large number of fragments, *and these are drawn at random from the manufacturer's bulk* he should expect the mean to be at least 27. This conclusion cannot be drawn by reading the B.P.C.

Weight per unit area. The B.P.C. does not specify the method to be employed in determing the weight per unit area. The Handbook gives two methods. One requires 3 portions, $6\frac{3}{4}$ in. square, to be cut from different parts of the cloth with the aid of a template, and the average calculated to ounces per sq. yd. The other depends on measuring the area of a weighed piece of cloth of any size or shape and calculating the result in the same terms.

The South African Bureau of Standards[4] approximately follows the template method of the Handbook, but in the case of lint specifies with precision the distribution of the 6 test pieces over the area of available cloth and suggests a 4-in. × 4-in. cutting die. In this laboratory, a 10 cm. × 10 cm. template is used—almost identical with the South African Bureau template, but giving results in metric measure directly.

As in the case of threads per inch, it would seem at first sight that these methods should all give the correct answer. There is no obvious anomaly, but the errors of determination have not been investigated. The different methods have, however, different statistical implications.

According to Davies[1] (p. 207) the relation between the size of the specimen and the variance in the case of solids, is not deducible from theory, but must be experimentally ascertained for each case. It seemed likely, *a priori*, that the variance of weights per unit area calculated from large pieces of cloth measured and weighed would be less than that of small template samples. If this were so, then the probability that any particular determination of the weight per unit area of a bulk quantity of cloth should lie within any particular distance of the mean weight would be different for the different methods. Preliminary checking of this provided confusing results. The co-efficient of variation of a large number of template samples taken from the bulk deliveries of bandage cloth in quantities of many tens of thousands of yards was 3·23 per cent. The co-efficient of variation of $2\frac{1}{2} \times 4$ yd. bandages cut from this cloth, weighed and measured, was 3·16 per cent.—almost the same figure. This result suggested that the variance was independent of the size of the sample, a result which could occur if all the variance was between the large bulk rolls and none of it was within rolls, so that the final effect would be similar to that found in sampling a bulk delivery of drums of liquid. From the nature of cloth, however, this reason seemed improbable and an experimental investigation was carried out.

The cloth chosen was unraised lint because it has a firmer texture than most surgical cloths and is less liable to distortion, but it is probable that similar conclusions apply to other cloths. A piece of cloth was taken and 289 squares of approximately 5 cm. sides were marked out, 17 in each direction, A number was given to each square. With a mechanically operated circular cutter 4·41 cm. in diameter, discs were cut out from each square and weighed (unconditioned) separately. From the known positions of each disc, combinations of any desired size could be made arithmetically. The whole of the cloth was not used since the discs were slightly smaller than the squares, but no serious error could arise from this point and the weight per unit area calculated from the disc weight has been assumed to be the same as the slightly larger square from which it was cut.

The detailed results are shown in the Appendix. Here it can be said that the true relation lies between that which would occur with uncorrelated discrete objects (coefficient of variation varying as $\sqrt{n}$) and the complete independence of sample size which occurs in the sampling of liquids. The size of the specimen and the method of calculating the weight per unit area are therefore not matters of indifference.

The analyses of variance (Appendix) show that in lint, the variance in the warp is a large and significant component of the total, whereas in bandage cloth, it is the weft which contributes most of the variance. This is to be expected from the proportions of the weight of the threads in the cloth in each case—lint having most of its weight in the warp, bandage cloth most of it in the weft.

The closeness of the variances of the bandages and of the small samples in the long series remains a puzzling feature. It could occur if the variance within rolls of cloth was very small compared with the variance between rolls, but the facts appear to be otherwise. In the case of lint the coefficients of variation of repeated samples from two different weavers was 1·76 and 2·45 respectively of the same order as that found by the disc method (2·2) for similar areas all drawn from a small piece of cloth. The most probable explanation is therefore that almost all the variance is within rolls so that any one roll can be regarded as containing all the sources of variation of the whole consignment, and the pattern of variation is such that the small specimens and the large ones are adequate measures of the same characteristic property. If for example, the main cause of the variation in bandage cloth is the change in weight of the weft yarn from bobbin to bobbin, then bandages which are not too long may vary from one another about as much as individual pieces from the bandages.

The trade average, Table VI, column 5, is 0·698, exactly on the B.P.C. minimum. As in the case of the threads per inch it is apparent that the meaning of "Weight not less than——" is that the bulk average is not less than this figure and the chance that any particular specimen exceeds this figure is about evens. In fact, 11 out of 19 specimens were over the B.P.C. minimum, a very fair approximation.

Statistical Implications

It now becomes apparent that the B.P.C. specifications are by themselves not capable of providing evidence as to whether or not a manufacturer is producing cloth to the specification if the analyst has only the book and a piece of cloth to guide him. The material is variable, and from a single specimen, the variance cannot be found and is not recorded. Without the variance he cannot say what the chance is that his specimen is representative of a bulk which is correct. He is not entitled to assume that every specimen in his hands will conform to the apparently rigid requirement "Not less than——." He is not much better off if he has several specimens, for estimates of variance are then subject to large errors and his specimens are quite likely to be correlated, since packages pass through the factory without necessarily becoming randomised. If he has sufficient experience, he may make mental allowances, but in that event he is really using statistical methods by intuition instead of by measurement. The U.S.P. variation of 3 threads per inch, with no statement of frequency, suffers from the same fault of leaving the variance unspecified. The magnitude of the possible errors is considerable as shown in the following paragraphs.

It is clear that co-efficients of variation differ widely according to the character under examination and range approximately from 1 to 5 per cent. What are the consequences to specifications and their interpretation?

As an illustration take the weight of open wove bandage. This is required to be not less than 100 grains per sq. ft. (0·698 g./100 sq. cm.). Is this the figure which results from the examination of a 6-in. × 6-in. square of cloth or from some other size such as a bandage? Is it necessary that the chance of drawing a light piece shall be so low that it "never" occurs? What is "never"—is it 1 in 10, 1 in 100 or 1 in 1000 or 1 in 10,000? Or is, as seems to be the case, 100 grains the average figure to which a large bulk of B.P.C. bandage reaches so that in sampling bulk made exactly to this average, failures to reach 100 grains will occur as often as in 50 per cent. of the samples? What is the effect of allowing a tolerance on the various figures so that small departures are disregarded?

When variation is very small, it may be possible to ignore it. For example, the limits for many pharmacopœial solutions have been set at ± 10 per cent., obviously an arbitrary number not derived from any painstaking measurement of the variation encountered in practice. But so long as the variation is much smaller than the limit no harm is done, because the object of the limit is merely the exclusion of gross errors, whether these are mistakes in manufacture, or fraud. If now, the same 10 per cent. limits are applied to material with a somewhat larger variation (say 3 per cent. co-efficient of variation) the limits become a means of controlling quality, for as many as 1 per cent. or so of the samples will fail, and this may cause trouble. Now apply the same limits to materials with a co-efficient of variation of 5 per cent. and we find that 1 in 20 samples will fail, quite enough to cause a lot of trouble. With a co-efficient of variation of 10 per cent., the limits become unworkable for a third of all samples fail so that penalties become absurd. The application of a fixed tolerance, such as 5 per cent. to properties with different natural variances has moreoever the effect of selecting the most variable property for the most criticism even when the variation is inevitable. In the case, for example, of gauze woven to an exact average of 19 × 15 with standard deviations of 0·42 and 0·73, the number of samples failing to pass within 5 per cent. for the warp will be far less (2·4 per cent.) than those for the weft (33 per cent.). There is no functional or logical reason for selecting the weft for a penalty 12 times as often as for the warp, once the natural variance is discovered.

A possible solution. If to each suitable figure in the specification, were simply appended the words "Standard deviation——," the difficulties would be much reduced. If in the preface or appendices, there were somewhat more detailed directions as to sampling and methods of examination the difficulties would be still lower. Provided that the variation is normal in the statistical sense a "population" of variable articles can always be completely specified by the mean and standard deviation and in fact this applies sufficiently well to properties which are not normally distributed. From the figures, different people can for their own different purposes, derive any information they want, as in the following examples.

The very large user or manufacturer can at once construct a "quality control" chart with ranges with which B.P.C. material should comply. A person examining a smaller but still considerable number of specimens can infer what proportion of them may be expected to fall outside any limits he decides to set. The person confronted with a single specimen can at least tell what the chances may be that his specimen, found to be faulty to a measured extent, is merely an extreme variant of bulk production which is actually correct. He will still have to decide whether say 1 in 100 is a small chance or not, but if he has a second specimen his confidence becomes greater, and is at all times expressible in figures which could indeed be tabulated for him.

There is actually no way, except "quality control," of avoiding all disputes between producer and consumer, but at least something can be done to narrow the field for disagreement. It is not necessary that a trade unused to such matters as fiducial limits and standard deviations should pay much attention to them, for the meaning of an average can be made apparent to all by slightly enlarging the definition and the manufacture by normal processes will not transgress the limits for standard deviation if these are set by a sufficiently wide view of the actual facts. The facts themselves either exist or can be gathered without much difficulty and their conversion to statistical form once and for all is a simple matter.

The confusion which undoubtedly exists is due (1) to an attempt to specify variable material without using a measure of its variation (2) to the widespread use of outdated methods of determination (3) to some lack of definition in terms which can be interpreted in more than one way. It follows from the proposal that the familiar range specification will disappear as a primary standard. There can be no objection to the retention of a range specification if this is correctly derived from the mean and standard deviation, but before this can be done, sampling procedure must also be specified. Single figure limits are even less desirable, for neither mean nor variance is then controlled and the absence of an *upper* limit *increases* the chance of extreme *low* variants being present in material which nevertheless complies with a specification merely based on an average.

In conclusion, these proposals, if accepted, would not change the dressings, nor require any modification of manufacturing methods. They simply consist of an accurate description of satisfactory practice available to and useable by anyone who cares to apply it to the goods before him. This cannot be done with the present specifications without the appearance of anomalies so large that they are repugnant to common sense. This is not because the specifications are individually wrong, but because it is impossible to construct a satisfactory specification for variable material not capable of being subjected to 100 per cent. inspection without recognising the fact of variation. Limits in such cases result in cost driving the properties in one direction, and the danger of being below the limit driving them in the other. The properties tend to become determined by commercial expediency and will be different for different producers and at different times according to a shifting balance of judgment. To some extent this is normal to every article entering trade, and as a matter of strict mathematics, there is no more difficulty in arranging standards and methods of inspection to cope with conditions in which 30 per cent. of the specimens fail to reach a standard than if the number was 3 per cent. or 0·3 per cent.

But it may be doubted if this is psychologically right. It has been stated as an empirical fact that continuously variable products with 1 in 200 defective according to specification often seem to satisfy producer and consumer. This level is very far indeed from some discussed in this paper, and the amount of explanation required to deal with the position seems excessive.

It must be emphasised that the discretion of the analyst, whether of producer or consumer, is not impaired by the use of exact specifications. In this paper, the phrase "significant difference" is used in a strict statistical sense, and merely means that the difference cannot reasonably result from chance. The word "reasonably" can, if necessary, be given numerical value. It does not at all follow that a significant difference is important, and the degree of action in the event of the discovery of a significant departure from specification remains in the discretion of those concerned. Perhaps this discretion can be best exercised if judgment receives exact information on significance before it is required to pronounce on the importance. It is quite certain that no valid pronouncement on importance can be made about differences which are not significant.

Summary

1. A number of the B.P.C. characters and tests for surgical dressings are examined in a statistical manner. It is shown that the satisfactory practice of the trade is imperfectly described by the existing specifications.

2. This is attributed to the lack of measures of variation satisfying statistical requirements, to the existence of trade customs not apparent from the specifications themselves and in one case to the use of traditional but ambiguous methods of determination, and in others to the employment of terms which on examination can mean different things to different people.

3. It is proposed that specifications in which the fact of variation is important should all have attached to them a measure of this variation. In general, the mean and standard deviation should be used. The numerical values for these should be fixed so that the existing dressings are described and not so as to compel modifications which on other grounds are not required. Where trade custom or the literature is ambiguous, definite references to, or descriptions of methods of examination should be more freely employed.

Appendix

Weft threads per inch, B.P.C. Bandage.

1. Analysis of variance of counts on the same piece of cloth by 3 people using 5 methods, e.g., the 10-in. count, the diffraction grating and the 1-in. counting glass methods of the Handbook, and two traditional ways of using the 1-in. counting glass. All the original figures were coded by subtracting 26.

The analysis of variance is shown in Table I.

Application of "t" to the differences between methods. Interaction must be used as the error variance because it is significant.

TABLE I

Analysis of variance	Sums of squares	Degrees of freedom, ϕ	Variance
Between methods	23·383	9	2·600
„ people	0·672	2	0·336
Interaction	1·653	3	0·551
Total between subgroups	25·708	14	
Within subgroups (= error)	20·142	135	0·149
Total	45·850	149	

Variance ratio: (F)		F	P
Interaction/error	$= \frac{\cdot 551}{\cdot 149}$	= 3·7	Nearly 1 per cent. point
Methods/error	$\frac{2\cdot 600}{\cdot 149}$	=17·5	Highly significant
People/error	$\frac{\cdot 336}{\cdot 149}$	= 2·25	Approx. 10 per cent.—not significant

Coded means of methods are: 10 in. = 1·363, B.S. = 1·690, Grating 1·340. Max. = 3·300, Min. 1·200.

Variance of mean 30 observations on each method is $\frac{0\cdot 551}{30} = 0\cdot 01839$

$$\text{difference of means} = 0\cdot 03678$$
$$\text{S.D.} = \sqrt{0\cdot 03678} = 0\cdot 192$$

5 per cent. value t for $\phi = 3$ is 3·18. A difference of $3\cdot 18 \times 0\cdot 192 = 0\cdot 61$ is needed for significance.

None of the official methods therefore differs significantly and neither does the minimum glass method. The one which does stand out as significantly different is the traditional maximum thread method.

t test applied to people

R.M.S. 74/50, F.W. 82·3/50, M.K. 80·5/50
= 1·48 = 1·64 = 1·61

$$\text{Variance of mean for 3 people} = \frac{551}{50} = 0\cdot 111$$

$$\text{S.D.} = \sqrt{0\cdot 0222} = 0\cdot 149 \quad \text{5 per cent. value for t } (\phi = 3) = 3\cdot 18$$

None significantly different.

The variances of the 5 methods have been calculated as: 10 in. count, 0·115; B.S. glass, 0·294; grating 0·217; traditional maximum 0·535; traditional minimum 0·48. (In considering these, or making detailed calculation, regard must be paid to the fact that the actual length of cloth counted in the methods is, 10 in., 10 in., 9 in., 1 in., 1 in.) The traditional methods are not so inaccurate as they appear, their main fault being ambiguity. The grating count has a standard deviation of 0·27 for a mean of 10 (= 30 separate readings) so that the mean is estimated with a standard error of 0·09. It is clear that even omitting all manufacturing error from piece to piece, variations of 0·3 threads per inch can easily arise as mere errors of determination and errors of sampling within a fragment of cloth, and the trade practice of disregarding small variations is mathematically justified. With large quantities of material available the mean can, of course, be estimated with any desired degree of accuracy, and ought then to agree with specification.

TABLE II

THE EFFECT ON THE COEFFICIENT OF VARIATION OF VARYING THE SIZE OF THE SPECIMEN IN MEASUREMENTS OF WEIGHT PER UNIT AREA

	Found in cloth	Calculated $\frac{\text{single}}{\sqrt{n}}$	Found by randomising the discs
Single discs..	2·83	2·83	2·83
Fours (2 × 2)	2·20	1·415	1·46
Nines (3 × 3)	1·67	0·943	0·936
Sixteens (4 × 4)	1·18	0·707	0·810
Twenty-fives (5 × 5)	0·98	0·566	0·508

It is clear that the results for the cloth as it lies are due to the correlation which exists between neighbouring areas in the cloth. When this is removed by randomisation the results agree well with a calculation based on the formula for samples of discrete uncorrelated objects.

TABLE III

THE VARIATION IN WEIGHT PER UNIT AREA OF BANDAGES (4 YD. × 2½ IN.) CUT FROM A SINGLE PIECE OF CLOTH

A roll of cloth was sampled as it passed through the production cutting and rolling machine, so that 36 bandages were taken in 6 lengths and 6 side by side. The piece of cloth sampled was therefore 24 yards × 15 inches. Each bandage was lettered and numbered to indicate its position, dried, conditioned weight calculated and then measured. Weight per unit area was then calculated. 700 was subtracted from each weight × 1000. The data and the analysis of variance are as follows:

	1	2	3	4	5	6	Total	Mean
A	6	8	8	3	9	23	57	9·50
B	24	24	31	21	29	21	150	25·00
C	19	29	24	45	21	22	160	26·66
D	8	10	25	11	16	14	84	14·00
E	14	16	24	0	6	25	85	14·16
F	8	11	21	25	16	21	10·2	17·00
Total	79	98	133	105	97	126	638	106·33
Mean	13·16	16·33	22·16	17·50	16·16	21·00		17·72

	Sum of squares	Degrees of freedom	Variance	Standard deviation
Rows	1365	5	273	16·5
Columns	334	5	67	8·2
Remainder	1368	25	55	
Total	3067	35	87·8	9·46

Coefficient of variation 1·30.
Variance ratio:
Rows/remainder = 4·86
1 per cent. point 3·86
Rows are highly significant. Clearly the columns are not significant.

It is clear that the weight per unit area of these bandages varied considerably according to which *length* of cloth they came from, but that the variation among bandages which lay side by side in the original cloth was unimportant.

The mean for the whole sample was 0·718, well above the B.P.C. limit of 0·698, and no bandage fell below 0·700. The cloth would undoubtedly have passed any official inspection. In spite of this, however, the tables of the normal deviate show that 1·5 per cent. of similar bandages would fail to reach the B.P.C. limit. The chance of this failure is not equal in different parts of the same roll of cloth.

A comparison of lint cloth from two different weavers

Samples taken over a long period gave the following results for the means:

	Warp	Weft	Weight
Maker 1	39·8	24·3	1·96
Maker 2	39·9	24·3	1·98

By all ordinary judgment, both weavers were producing correct cloth of identical characteristics. When, however, the variances were examined, it was found that for the weft, Maker 1's variance was 0.087, but Maker 2's was 0·344. These are significantly different at the 0·01 level. Maker 2's machines were evidently not producing such even results. The practical importance is that the chance of any individual specimen being below standard is much greater with Maker 2 in spite of the identical averages. A specification based on average alone is incapable of dealing with this situation.

TABLE IV

WEFT THREADS PER INCH FROM 23 CONSECUTIVE DELIVERIES OF BANDAGE CLOTH

Weft threads per inch. 27·0, 27·8, 27·2, 27·7, 27·0, 27·6, 27·2, 28·0, 27·4, 27·2, 27·6, 27·8, 27·8, 27·0, 27·8, 27·0, 27·9, 27·0, 28·4, 27·3, 27·2, 27·7, 27·7.

Standard deviation 0·48. Mean 27·49.

The excess needed to comply with specification

Assume that bandage cloth has a mean weight of 0·800 and a standard deviation of 0·07. Then 7·2 per cent. of 100 sq. cm. specimens will be below the B.P.C. limit of 0·698, and nearly 3 per cent. will fail to pass B.P.C. limit—5 per cent. To ensure that as few as 1 per cent. of all specimens fail the B.P.C. limit, the average must be 23 per cent. over the limit.

For B.P.C. gauze, coefficients of variation of 2·19 per cent. and 5·16 per cent. for warp and weft threads per inch, a manufacturer would have to provide 19·98 and 16·72 threads per inch if he is to have as few as 1 per cent. "defective" specimens.

TABLE V

WEIGHTS OF 27 CONSECUTIVE 100 SQ. CM. PIECES CUT FROM A SINGLE BANDAGE

The weights in mg. less 700 are given in order as follows:

176, 184, 225, 229, 177, 157, 171, 134, 160, 218, 239, 239, 99, 72, 105, 95, 45, 73, 98, 88, 52, 99, 104, 107, 77, 46, 38.

Mean: 0·830. Coefficient of variation 4·23 per cent.

This case is interesting for it shows clearly (*a*) the sudden change in weight at the 13th reading, evidently due to a change in the weft yarn at that place, (*b*) the fact that although the bandage was variable, the whole bandage and every fragment conformed to the B.P.C., (*c*) that this conformity was only due to the high mean. With a more normal mean and the same variance, parts would have been below standard.

TABLE VI

CHARACTERISTICS OF SOME SURGICAL DRESSINGS PURCHASED BY RETAIL AT RANDOM
2 in. × 4 yd. Bandages, April, 1951

Number	Length	Width	Weight	Weight per 100 sq. cm.	Warp	Weft
6033	3/30″	2″	12·4	0·697	47·0	27·8
3968	3/34″	,,	12·5	0·684	45·0	26·3
6004	4/1″	,,	12·1	0·647	44·0	25·7
6002	3/34″	,,	12·6	0·689	48·0	27·1
6023	3/35″	1·15/16″	11·8	0·660	45·0	26·8
6025	4/6″	2″	13·3	0·688	45·0	28·0
3960	4/5″	,,	13·8	0·719	42·0	28·0
3964	4/8″	,,	13·0	0·672	44·0	26·0
3961	3/33″	,,	13·0	0·715	46·0	27·0
3963	3/35″	,,	13·4	0·726	48·0	25·3
3962	4/2″	,,	13·2	0·702	45·0	26·8
6008	3/33″	1·7/8″	12·2	0·730	46·0	27·0
6024	4/0″	2″	13·5	0·726	48·0	27·0
3976	3/34″	,,	12·4	0·676	44·0	26·2
6042	3/32″	,,	12·1	0·670	48·0	28·1
6044	3/33″	,,	13·6	0·749	48·0	28·5
6046	3/34″	,,	13·5	0·737	44·0	28·4
6045	4/1″	,,	14·2	0·759	47·0	26·7
6005	3/26″	,,	13·5	0·781	50·0	27·4

Extremes: 92·6–111·8 per cent. of the minimum B.P.C. weight. Range: 19·2 per cent. 8 out of 19 are below B.P.C. weight.

TABLE VIII

B.P.C. GAUZE SURVEY, MAY, 1951
1 yd. unless otherwise stated

Origin	Length	Warp	Weft	Weight	Comments
6206	35	20·0	16·0	13·0	
6189	36	19·6	15·2	13·0	
6207	36	19·0	15·0	13·3	
6208	36	21·0	15·0	12·0	
6185	35½	19·4	15·0	12·2	
6186	17	28·5	24·6	14·3	(½ yd.) This gauze is B.P.C. but the construction is unusual, being a closer material.
6170	36	20·4	14·1	12·5	
6166	35	19·9	15·6	12·9	
6180	36	19·6	14·2	12·2	
6201	36	19·4	13·7	12·2	
6204	36	19·0	14·5	12·4	
6196	36	19·2	14·6	13·6	
6219	36	19·9	14·6	11·8	
6155	33	19·6	15·0	12·4	
6218	39	20·0	14·8	11·8	
6167	36	19·5	14·5	12·1	
6168	36	19·7	14·0	12·7	
6209	34½	20·4	14·2	12·0	
6210	36	19·0	15·3	13·2	
6217	36	19·0	14·4	12·7	

The term 4 yd. can have the following meanings.

(1) Bulk average of production 4·00.... yards. Standard deviation unknown and undefined. Length of any single bandage therefore uncertain. 50 per cent. of short bandages.

(2) "No" bandage less than 4 yards. The word "no" has no definite meaning when applied to variable material not subject to 100 per cent. inspection. It never really means that the chance of a short bandage occurring is zero. The choice of the mean depends on the "higgling of

the market" and both buyer and seller commonly delude themselves until each is more or less satisfied that "no" bandage is less than 4 yards.

(3) "4 yards of bandage," in the sense of the Weights and Measures Act. This requires that 100 per cent. inspection should be applied, a virtual impossibility for this article cut and rolled in a single operation.

I acknowledge with thanks discussions with a number of people, particularly Messrs. A. W. Evans, L. H. C. Tippett, D. M. Bryce and A. D. Rhodes, and technical assistance from Misses F. Wilbraham and M. Kennedy. For the views expressed in the paper I am alone responsible.

References

1. Davies, *Statistical Methods in Research and Production.* 2nd Ed. Oliver & Boyd, London, 1949.
2. *Methods of Test for Textiles, B.S. Handbook No.* 11, Ed. British Standards Institution, London, 1949.
3. *Shirley Institute Test Leaflet A.C.5.* British Cotton Industry Research Association, Manchester.
4. South African Bureau of Standards, Pretoria.

DISCUSSION

The paper was presented by THE AUTHOR.

The CHAIRMAN suggested that the new concept which Dr. Savage had introduced would be carefully studied by those responsible for fixing standards for dressings.

MR. J. R. ELLIOTT (London) compared the testing of bandages, in which only one was used, with testing a batch of tablets for disintegration when a number of tablets were taken and if these failed the test the test was repeated before the whole batch was rejected. The bandage was only one small section of a piece of cloth. Should not a larger number of samples be taken before rejection?

DR. K. R. CAPPER (London) said the analogy with tablets was not quite accurate because the analyst examined a number of pieces of a dressing. The phrase "not less than" when applied to the number of threads per inch in a bandage assumed that the manufacturer would so prepare the material that, taking variations into account, it could be expected to contain that minimum number with a reasonable probability. He did not know how, in practice, a standard such as "the count must be so-and-so plus or minus two or three" could be applied. He had no doubt Dr. Savage's proposals would be considered by the Surgical Dressings Sub-Committee of the B.P.C.

DR. W. MITCHELL (London) suggested that if one bandage was specified for a patient it was in order to test that one bandage to see whether it was suitable for the purpose required.

MR. R. L. STEPHENS (Brighton) pointed out that the attitude of mind of tablet makers and cloth makers differed widely.

DR. SAVAGE, in reply, said that he had examined the B.P. limits for tablets, making certain assumptions, to find out what was intended to be the standard deviation of a single tablet. It was of the same order as with a surgical dressing. There was no particular difficulty about analysis of surgical dressings, which could be treated in the same way as anything else. He described American practice in controlling the quality of surgical dressings, which ensured that no action was taken until the material in bulk had been proved faulty. That was different from the procedure in this country but it corresponded more nearly to his views. He had collected evidence to show that the trade worked to a bulk average, i.e., the average of the production. The manufacturer and the pharmacist looked at the problem from two different points of view. The pharmacist expected every sample to comply. When the manufacturer replied, "But my material does comply," he meant what he said—but the two people were not talking about the same thing. Dr. Savage said he wished to dispel the belief that the problem could be settled if the standard were raised so that every specimen complied. Strictly statistically, there was no such thing as "always" or "never"; it was purely a question of measuring probability. Once the cloth had been made it could not be altered. A certain amount of variation must be accepted with all variable products. They should recognise it and put a mathematical standard to it. The variation which occurred between normal surgical dressings was so much smaller than that which would be detected in practical use that it was not of great importance except from the analytical point of view. Nevertheless, the user was entitled to protection, and his proposals would provide this.