JOURNAL OF Pharmaceutical Sciences

January 1966 volume 55, number 1

Review Article

Application of Statistical Methodology in Quality Control Functions of the Pharmaceutical Industry

A Survey

By T. N. T. OLSON and I. LEE

CONTENTS

INTRODUCTION

"OF WHAT Significance Statistics" was the title of an editorial by Edward G. Feldmann in the November 1963 issue of the *Journal of Pharmaceutical Sciences* (1). Feldmann recognized the increased use of statistics and commented on the merit of the uses to which statistics were then being applied. This paper, which is limited to the application of statistical methods to pharmaceutical quality control, is intended to amplify that editorial.

There has been a steady increase in the use of statistics in the quality control functions of most of the companies in the pharmaceutical industry over the past 10 years. This is not too well evidenced when examining the literature of the pharmaceutical sciences, but is readily apparent if one examines the technical statistical literature.

The past few years have shown that the analytical controls, and those after-the-fact subjective or inspection-type controls employed during the manufacture of a pharmaceutical product may be about only 10% of the effort used to control a product adequately. The other 90% would consist of the people, material, facilities, building, equipment, product design, procedures and specifications, and documentation under which and by which the product is manufactured.

Statistical methods applied to outgoing material, and control charts on analytical assays are after the fact, that is, after the material or product has been manufactured. Statistical methods applied to filling operations or on-line inspection of products are of much more value since they are controlling the material at the time of manufacture. The 1962 drug amendments to the Food and Drug Act of 1938 acknowledge that it is the in-process controls applied to a product which in effect make a quality product.

Five years ago statistical sampling plans with a sensitivity in the range of 1 to 5% defective

Received from the Quality Control Division, Parke, Davis and Co., Detroit, Mich.

were considered adequate for inspection of packaging supplies. Now engineers would like 0.1% sensitivity for certain types of defects which interrupt high-speed packaging lines. When problems such as these arise, statistical methods must be modified for a different approach to the problem because the cost of sampling 3000 to 5000 units would be prohibitive, let alone the cost to manufacture certain materials with a value of defectiveness so low.

Mainland (2) poses 2 interesting questions that could be applied to our desire for increased knowledge and perfection: "If the verdict of a test is 'significant' at our predetermined level, what will we do in consequence thereof?" And the second question: "If the verdict of a test is 'not significant' at our predetermined level, what will we do in consequence thereof?" These two questions applied to a series of alternative quality control situations will soon show that the way to produce a product is to design it right and do it right the first time. It is obvious that sampling plans which examine 10 out of a million tablets do not adequately control a product. A sampling plan which weighs 1 out of each 100 tablets as they are produced is controlling the product.

The statistical methods which are discussed in this paper are not discussed from an "after-the-fact" or "during-processing" concept, but it is obvious that the value of statistical methods will be the greatest if directed at the during-processing concept. Statisticians working in the quality control field all find quite a change from dealing with the research aspects of pharmaceuticals. The scientist who is inherently an optimist may accept odds of 5 to 1, the statistician being analytically a pessimist prefers odds of 20 to 1, but the quality control man wants zero defects, unrealistic as this may be.

STATISTICS OF ANALYTICAL PROCEDURES

The accuracy of an analytical procedure can be established by applying it to a product or material containing known amounts of the material to be assayed or by using unknowns and running both the new procedure and another whose accuracy has been established.

Accuracy refers to the closeness of the result to the true chemical content. Precision refers to the agreement of repetitive results with each other (3). The accuracy of an analytical test would be the mean of any repeated results as compared to the amount of material predicted to be in the product. The standard deviation may be estimated from the repetitive assays performed on the product and is a measure of the precision.

Youden (3) describes a simple method of analyzing analytical data by plotting the amount found against the amount taken for analysis. The plot should be a straight line. Figure 1 indicates the types of problems encountered with assays. A constant per cent bias of the slope is an inability to obtain in the amount found the same per cent increase that was in the amount taken. A constant intercept bias is the inability to recover the amount found in proportion to the amount taken.

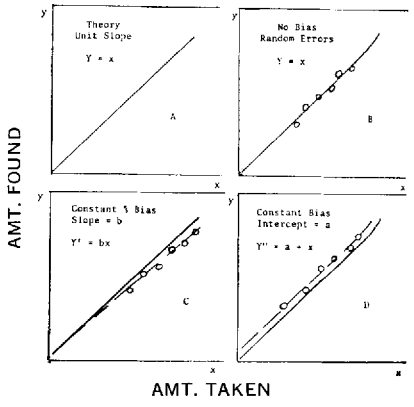

Fig. 1.—Diagram to show the effect of random errors, constant per cent errors, and constant error of analytical results. [Figure 1 of Youden (3).]

Youden describes the many types of problems encountered with the evaluation of new analytical methods including comments on Round Robins. He suggests a half-dozen different materials might be sent to a minimum of 5 different laboratories with requests for only single analytical determinations. He asserts that it is not necessary to ascertain the laboratory's precision because of the large number of results necessary to estimate the precision, the usual faithfulness of deviations from a procedure, and the possibility of censoring the duplicated data.

In a series of papers (4–8), Garrett *et al.* discuss the selection, evaluation, and control of the assay of several pharmaceutical products. Garrett makes use of many statistical techniques in these papers, including control charts, analysis of variance, design of experiments, and degradation rate statistics in addition to the usual statistical tests of significance. Statistical treatment of the data permitted Garrett to determine optimum solvent quantities for separation of the active ingredient, the number of dosage units to assay within a desired confidence limit, the reproducibility of fill weights, the homogeneity of the fill mix, and other relationships between assay economy and product variability. This series of articles provides a good summary of the statistical

methods available to an analytical laboratory in evaluating assay and dosage form variation.

Davies (9) outlines an economical testing program whereby the costs of routine analytical tests are compared to the costs of accepting material which is either above or below stated specifications. His cost for the rejection of acceptable batches (the α-error) is on the basis of reprocessing cost; the cost for accepting unacceptable batches (the β-error) is not quite as easy to quantify.

In the pharmaceutical industry, the probability of wrongly classifying good material is rather low because of the tendency to attempt to test quality into the product. It would be expected that any material to be rejected would be tested sufficiently to prove that point; and so the cost would only be for additional testing. These points are explored further in the paper.

Calder (10) suggests a revised method of calibrating instrumentation relating to spectrochemical analysis where the calibration procedure involves plotting the instrumental response for an element against the concentration of that element. By calculating regression lines using the method of least squares for the response *versus* concentration, only 24 instead of 84 spectra and only 18 hr. instead of 60 hr. were required as compared to the original method.

Wernimont (11) used a computer to compare 16 Beckman spectrophotometers as to (*a*) the variation of the absorbance curve of each spectrophotometer about its own absorbance curve, (*b*) a comparison of the variation of a selected group of spectrophotometers about their average absorbance curve, and (*c*) a comparison of the remaining unselected group of spectrophotometers to the standards and tolerances of the selected group. Spectral analysis was used to analyze the data. The data were explained by the computation of 2 characteristic vectors. The first vector related to the vertical displacement of the absorbance curves and the second vector to the horizontal shift of the curves. Wernimont described a tentative absorbance standard of 60 mg. potassium chromate per liter, together with a procedure to check the Beckman spectrophotometers and a high-speed computer program which will make the necessary transformation to scale multiples of the 2 vectors.

The current growth in the use of statistics as applied to analytical chemistry can be attributed to an increased awareness of the colleges to mathematics in general. The use of automated analyses and computers will ultimately require extensive use of statistical methodology in the modern analytical laboratories.

SAMPLING PLANS

Description of Sampling Plans.—Acceptance sampling, which received its start during World War II, when it was used to pass on the quality of armaments, has been accepted for a large number of uses by the pharmaceutical industry.

An acceptance sampling plan may be described as follows. A company produces a lot of compressed tablets. It obtains a representative sample of tablets from the lot, examines the sample and, based on information obtained from the sample, either accepts the lot as conforming to standards or rejects it. The sampling plan is defined by (*a*) the size of the sample taken and (*b*) the number of defectives or degree of defectiveness allowed in the tablets. Graphically, a sampling plan can be explained by an operating characteristic curve.

Four points are defined on the operating characteristic curve selected for an example by Breunig and King (Fig. 2, *A*). First, a lot of tablets is said, for the sake of illustration, to be satisfactory if it contains only 10% defective tablets (the acceptable quality level, AQL). Second, a lot of tablets is said to be unacceptable if it

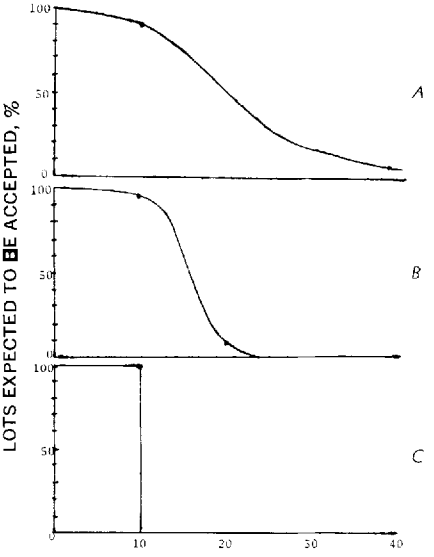

Fig. 2.—Key: *A*, operating characteristic (OC) curve for tablet example from *Reference 15*, p. 11, illustrating AQL = 10%, UQL = 40%, R_p = 10%; R_c = 8%, n = 10 [figure 5 of Breunig and King (12)]; *B*, typical operating characteristic curve from *Reference 15*, p. 27, illustrating AQL = 10%, UQL = 20%, R_p = 5%, R_c = 10%, n = 85 [figure 4 of Breunig and King (12)]; *C*, ideal operating characteristic curve illustrating perfect discrimination but unrealistic stringency [figure 3 of Breunig and King (12)].

contains more than 40% defective tablets (the unacceptable quality level, UQL). Third, there is a certain probability that a lot which in reality is 10% or less defective is observed on test to be more than that with the result that the lot is wrongly rejected. This is the producer's risk or α error. Fourth, there is the probability that a lot which is truly 40% defective is observed on test to be less than that with the result that the lot is accepted. This is the consumer's risk or β error. Once the above four values are selected the sample size is automatically determined.

Figure 2, *B*, shows an operating characteristic curve where the UQL has been decreased to 20% with a sample size of 85 tablets, and Fig. 2, *C*, shows a curve where the AQL and UQL are identical. To obtain the curve shown in Fig. 2, *C*, the entire lot of tablets would have to be examined.

The ability of a sampling plan to discriminate is dependent upon the size of the sample. Sampling plans requiring 10 to 30 samples have relatively low powers of discrimination.

There are two types of sampling plans: attribute and variables. Attribute sampling refers to a zero-one situation where a tablet is either good or bad; variables sampling is based on a continuous distribution of degrees of defectiveness and covers the gray zone between good and bad situations.

Attribute sampling plans require only the counting of the number of defectives found in the sample, or a mean or percentage of defectiveness [100 × (number of defects/sample size)]. The acceptance or rejectance of the lot depends on whether this value is smaller or larger than the one stated by the sampling plan.

Variables sampling plans require the calculation of a mean and standard deviation or range. For example, a sample of 10 tablets is taken which should have a theoretical mean weight of 100 mg. The mean weight of the sample is 94 mg./tablet, with a calculated standard deviation of 2 mg. Using the formula (100 mg. − 94 mg.)/2 mg. = 3.0, and entering a set of tables similar to the normal tables, it is found that the value of 3.0 indicates the material is 1.2% defective, which is acceptable or unacceptable depending on the values of acceptance.

Double sampling refers to the following example situation. A sample of 50 tablets is taken. If it contains 2 or less defectives, the lot is accepted. If it contains more than 4 defectives, it is rejected. If it contains more than 2 defectives but less than 4 defectives, a second sample of size 50 is taken. If in the 2 samples, there are less than 4 defectives, the lot is accepted. Triple and multiple sampling are merely extensions of the above premise.

Sequential analysis of sequential sampling is an extension of multiple sampling developed by Wald during World War II (13). Successive samples are taken based on a varying set of criteria until either a decision to accept or reject is made. There is, of course, the possibility that no decision could ever be made and the samples would stay in the indifference zone indefinitely. The problem is easily resolved, however, by deciding to stop at a given sample size.

For practical purposes the need to design a sampling plan has been eliminated by a series of government-sponsored sampling plans, 2 of which are MIL-STD-105D for attribute single, double, and multiple sampling plans (14); and MIL-STD-414 for variables sampling plans (15). These books have gained acceptance throughout most of United States industry in a manner much like the U.S.P. and N.F. Government contracts for the purchase of pharmaceuticals usually refer to one or both of these books. The obvious advantage of selecting plans from either of these books is communicability and acceptance throughout industry. Hence, there is little or no advantage to specially designed sampling plans

Application of Sampling Plans.—Pharmaceutical products which are solutions are sampled ordinarily with a sample size of 1. This is based on the premise that a solution is a homogeneous mixture in which every milliliter is like every other milliliter. Suspensions are frequently sampled the same way, but this is predicated on a thorough or continuous mixing. It is frequently necessary to take more than 1 sample from a suspension to verify that it is truly homogeneous. If a lack of homogeneity is known to be present, then a suspension could be thought of statistically as resembling a lot of compressed tablets with an infinite population size.

If a drum of a powdered chemical is known to be homogeneous, then for statistical purposes it can be thought of as a solution. Single or duplicate samples then are considered sufficient to obtain a reliable response by an analytical test. If a drum of chemical is known to come from a process where the entire lot of chemical is not blended prior to being filled into drums, there can be no assumption of homogeneity for certain properties of the chemical. For practical purposes an infinite population size could be assumed such as for a lot of tablets. The reliability of 1 or 2 samples from such a drum would be of serious doubt no matter how good the

answers appeared. According to MIL-STD-105D, sample sizes in the range of 2500 may be necessary for attributes, and for MIL-STD-414, sample sizes of 200 may be necessary to obtain adequate representation of an entire production lot. It is obvious that if sample sizes of this size were used, they would be economically impossible to test. The military plans, however, allow the dividing of the production lot under certain conditions into inspection lots (subsets of the production lot) which substantially reduce the astronomic sample sizes mentioned above.

A situation analogous to the sampling of pharmaceuticals in the powder state was studied extensively by Duncan (16–18) under the auspices of the National Plant Food Institute. Four different fertilizers, 3 different sampling instruments, and 3 different laboratories were used in the experiments. It was noted that under certain conditions, 1 of the instruments showed a tendency to take in a higher percentage of larger particles and a lower percentage of smaller particles, which in turn gave a higher assay value. Little or no evidence was found to indicate that instrument sampling differed more on the average than samples obtained by riffling. Relative differences between laboratories were noted along with differences between days from the same laboratory. The official method of the Association of Official Agricultural Chemists for sampling bulk material requires a sample from 10 bags for lot sizes greater than 10 bags. One tube core is removed from each bag. For lot sizes less than 10 bags, at least 10 cores are to be taken, but at least 1 core from each bag. From a blended sample of the 10 cores, a sufficient quantity is taken for analysis. Duncan assumes the running of 2 tests on the samples, no more, no less, in his mathematical model of the sampling and assay procedure.

The work done by Duncan in the fertilizer experiment would lend itself to the formation of an official method of sampling bulk pharmaceutical powders where the mean of the replicate assays is the standard for acceptance.

Duncan (19) devised the operating characteristic curves for fertilizer inspection plans which utilized sample sizes of 20 and 10. Ten samples were suggested for use because of the lower cost of sampling which is comparable to the pharmaceutical problem.

All of the above discussion is useful as long as it can be assumed that the characteristics being sampled are randomly distributed between bags or packages. However, the writer has investigated the problem of powder in barrels from a lot containing varying particle size distributions, although the material was accepted on the basis of its chemical purity. In this case, particle size distribution in certain ranges caused undesirable physical properties in the final product.

For adequate sampling in a situation such as this, reversion to knowledge of how the powder was manufactured is a necessity if homogeneity of a property is known to be a problem. Statistical sampling of the product is no substitute for manufacturing the product by a process known to produce satisfactory characteristics. All that should be necessary is to run an identity test of the material.

Literature on appearance of pharmaceutical dosage forms is rather conspicuous by its absence in the case of sterile types of products. As a result, most manufacturers inspect 100% of all sterile products for appearance-type defects such as particulate matter within the vial. It apparently is not desirable to say that a given lot of ampuls contains particulate matter in 1% of the vials. Still, at the present state of the art, particle-free sterile products are virtually impossible to manufacture and also exceedingly difficult to inspect on an economic basis. Each company, however, must have a set of standards and a method for verifying that the standards are being followed. This represents its level of excellence. Attribute sampling techniques can do this verifying with ease.

Tablets and capsules are frequently 100% inspected. Tablets, however, frequently are produced with a very high degree of excellence and also do not carry with them the problem of sterility and injectable elegance. Sampling plans therefore, work well with tablets, except for the large lot sizes which, depending on how the lot is sampled or divided, may require that large sample sizes be taken.

If tablets are inspected as they come off the compressor and little or no capping is noticed, it is not uncommon to find an unacceptable level of capping a few days later when they are packaged. This, in reality, is saying that a measurement for the potential to cap has not been employed in the inspection process; but in addition it says that the tablets were not manufactured correctly the first time.

What is the proper way to handle a tablet containing a metal chip traced to the feed frame of the compressor? Is the whole lot to be 100% inspected, a portion of the lot 100% inspected, or more samples taken? This is not really a problem in statistics, but in quality control.

The statistics of weight variation have been undergoing a gradual but continuous refinement in the pharmaceutical industry. The following

discussion outlines the various approaches that have been investigated over the past two decades along with the problems associated with the various plans.

The U.S.P. XIV and N.F. IX had the following weight variation test for compressed tablets (20):

"Weigh 20 whole and uncoated tablets and calculate the average weight. When weighed singly, not more than two of the tablets deviate from the average weight by a greater percentage than that shown in the following table, and no tablet deviates by more than double that percentage."

Average Weight	Percentage of Deviation
13 mg. or less	15
More than 13 mg. and including 130 mg.	10
More than 130 mg. and including 324 mg.	7.5
More than 324 mg	5

(The above table has been modified in U.S.P. XVII by eliminating the first weight class.)

Dunnett and Crisafio (21) derived the operating characteristic curves for the above official tablet weight variation method using sample sizes of 10, 20, 50, and 100. It was found that by simulating a batch containing 5% defective tablets (too heavy or too light) the lot would be accepted 93, 95, 98, or 99%, depending on whether a sample size of 10, 20, 50, or 100 was used. A batch containing 20% defective would be accepted, 40, 23, 4, or 0% using the same sample sizes. It was concluded that a sample size of 10 tablets had only meager ability of protection against inferior products. The use of a standard deviation test for 10 and 20 tablets was discussed and also a 2-sample attribute plan requiring a total of 50 tablets. If the first sample of 20 was satisfactory, the remaining were not examined. The operating curve for this was better than the operating characteristic curve for the 10- and 20-tablet standard deviation plans.

Non-normality of tablet weight variation was studied. The higher proportion of non-normal lots found, than was expected, was attributed to (*a*) the incapability of machines to turn out uniform tablets, (*b*) the differences between punches, (*c*) the sampling at various times from the machine. Dunnett concluded that no reasons could be found for the substitution of another mathematical distribution in place of the normal distribution.

Smith (22) suggested as an alternative to the B.P. and U.S.P. methods of weight variation the use of sequential analysis using formulas given by Wald (13). The use of half-defective (tablets deviating by half the specified amount) and double-defective (tablets deviating by double the specified amount) were discussed, and a table for acceptance or rejectance using sequential analysis for half-defective tablets was presented. The advantage of using sequential analysis was in the fewer weighings (about half the number requested by the B.P.) required for a decision on acceptability. Disadvantages cited were the very small sample size which is to represent the lot and the ease of giving exactly 20 tablets to an analyst for weighing.

Green and Knudsen (23) discuss three types of sampling which were applied to samples of 20 dry-filled ampuls. These could be stated as (*a*) the average of the sample shall not be greater than or less than ___, and no individual ampul from the sample shall be greater than or less than ___; (*b*) the average of the sample shall not be greater than or less than ___, (*c*) and no individual ampul from the sample shall be greater than or less than ___. A comparison of the operating characteristic curves for the 3 plans indicates that the average plan, and the combined average and individuals plan are comparable and that both of these are better than the plan dealing with individuals. By analyzing collected data, it was asserted that a plan then suggested by the manufacturers in 1950 and a Canadian tolerance were too tight in the lower weight ranges and too loose in the upper weight ranges.

Breunig and King (12) described the advantages of variables sampling plans as having a greater ability to detect excessive tablet weight variation than the plan employed by the U.S.P. When Breunig and King compared their variables plan to a 2-step attribute sampling plan such as is now in the U.S.P. XVII and N.F. XII for assay variation, they found the variables plan superior for (*a*) cases of excessive variation and (*b*) for excessive variation combined with a shift in the mean and about equal for cases of mean shift alone. Two nonstatistical points were stated by Breunig and King which illustrate how inadequate the use of a statistical sampling plan is without simultaneous human thought. (*a*) A lot is accepted or rejected on the basis of an exceedingly small number of units (10 or 30 out of 1,000,000). As a result, it is not practical to expect a sampling plan to guarantee that all accepted lots will be of acceptable quality, although such a plan may be quite effective in detecting an occasional bad lot when all others are of good quality. (*b*) If many highly defective lots are submitted to such a plan, it is possible that half of them will end up being used by the consumer.

The solution to the above is an adequate quality control system throughout all the various stages of the processing of the product, which assures that the product is made right the first

time. The U.S.P. and N.F. sampling plans then become merely guide rules as to what level of quality is expected of the manufacturer of a pharmaceutical. From an enforcement viewpoint, however, the sampling plan's quality standards must be enforced on a lot-to-lot basis.

Haynes (24) demonstrated by the use of actual data from sterile solid weight variation and by computer simulation that certain suggested attribute plans were more robust than equivalent variables plans. Robustness refers to the problem which causes variables plans to be more susceptible to the effects of non-normal distributions commonly found in small samples (10 to 30 items). Figure 3, *A*, shows a normal distribution of weights as generated by simulation on an IBM 7070 with approximately 5% of the values outside the vertical dotted lines; Fig. 3, *B*, shows a platykurtotic distribution of weights; and Fig. 3, *C*, shows a leptokurtotic distribution of weights.

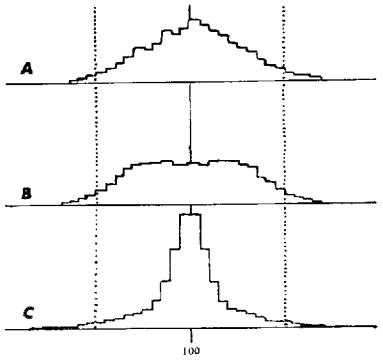

Fig. 3.—Key: *A*, normal weight distribution, quality index 94.5%; *B*, platykurtotic weight distribution, quality index 95.1%; *C*, leptokurtotic weight distribution, quality index 95.0%. [Figures 3, 4, and 5 of Haynes (24).]

Both platykurtosis and leptokurtosis are types of non-normality found in pharmaceutical dosage form weight and assay variation. Paul (25), by using data accumulated for individually assayed compressed tablets, reinforced the work of Haynes that pharmaceuticals were indeed manufactured in non-normal distributions as evidenced by samples of 10, 20, and 30.

In response to a request from the Food and Drug Administration, the Pharmaceutical Manufacturers Association, Quality Control Section, in 1961 set about to collect data for dosage form variation for sterile solids without diluents and for compressed tablets.

By October 1962, two differing plans were advanced by industry statisticians. These were the variables type sampling plan advanced by Breunig and King for compressed tablets and the attribute plan advanced by Haynes.

Although one type of plan was advanced for tablets and another for sterile solids, this was the result of two different committees working on the problem separately. Both tablets and sterile solids were, however, later considered using a single type of sampling plan.

Based on the work of Haynes and other industry statisticians, the P.M.A. Quality Control Section through its subcommittees recommended to the U.S.P. that an attribute type of plan be adopted in the U.S.P. for a selected group of tablets as a test for composition variation.

As a result, the U.S.P. XVII and N.F. XII have adopted the following statement regarding "content uniformity" for certain compressed tablet monographs: "Select a sample of 30 tablets. Assay ten of these individually as directed. The requirements of the test are met if all ten results fall within the limits of 85% and 115% of the average of the tolerances specified in the monograph. If one result falls outside these limits, assay the remaining 20 tablets individually. The requirements are met if not more than one of the 30 results is outside of the limits of 85% and 115%."

An interesting question arises when one asks, "What type of distribution of doses does the consumer have a right to expect when in fact he cannot be given on a repetitive basis exactly 5 gr. of aspirin?" Clearly, a leptokurtotic distribution gives the consumer on the average some of what he expects; likewise, a platykurtotic distribution of values shows a lack of control in a manufacturing process. If a normal distribution or leptokurtotic distribution is suitable for a customer, then is it not consistent to set the bias of the sampling plan against the platykurtotic distribution generated by a poorly controlled process?

Moskalyk *et al.* (26) expressed the observation that dosage variation was greatest in the lightest weight tablets within a batch. Then the question arose concerning whether this could be proved true for all tablets; if it was, a sample of lightweight tablets could be adopted as a control over uniformity of drug dosage.

Haynes (27) also was interested in control of weight variation as a means of controlling drug dosage.

It is clear that what is needed is a non-analytical test which statistically predicts the degree of potency variation between dosage forms. With the advent of the Mettler and Cahn automatic weighing balances, it is not impossible to weigh thousands of tablets, either as they are being processed or as a sample obtained from the finished lot. From the distribution of values obtained from them, a statistical test of normality

could be made. If a statistical technique can then be derived that would allow us to draw conclusions from 3 singly assayed tablets, the 3 lightest, or the heaviest, middle, and lightweight from a random sample of specified size, is it then not possible to predict with an acceptable degree of reliability that the lot conforms to U.S.P. or N.F. standards? Here is an area for some statistical research with great economic rewards.

Although attribute sampling plans are the simplest to use and simplest to enforce legally, variables sampling plans are still the most useful for the internal operations of a pharmaceutical manufacturer. Especially will this be true with the automation of the quality control function of pharmaceutical product manufacture.

The sampling of product container fills by a regulatory agency is properly a sampling technique, although it is routinely controlled by control charts and might logically be discussed under *Control Charts.*

The Department of Agriculture, State of California (28), has now put into effect Article 5 which will attempt to control from a legal viewpoint the concept of filling variation. The size of the sample of consumer size packages required for sampling follows closely the approach used by the Association of Official Agricultural Chemists for sampling bulk material in lot sizes of 10 containers or less. For lot sizes greater than 10 packages, the sample size approximates the square root of the number of packages in the lot. The plan, although not similar to either an attribute sampling plan or a variables plan, contains features of both type plans. The basic feature of Article 5 is that tables are presented for all possible situations which the investigator might meet. Article 5 does not say, however, for example, that a 5-lb. package must contain at least 95% of label claim. It merely says that for a given sample of package weights or fills, the average range and the average error are to be determined. Based on these values, unreasonable weight errors are to be determined, the values of which will determine whether the lot is acceptable or unacceptable.

True statistical sampling has very limited if any application at all in detecting a single foreign label or carton in a lot of labels or cartons. Only 100% nonhuman inspection of the entire lot will give any degree of reliability in detecting a foreign label or carton. Human inspection is second in preference to machine inspection. At the present time, it is better to consider machine inspection as a means of determining the presence of foreign labels, but not as a means of removing them from a lot of labels. It would be preferable to have the inspection machine stop rather than try to remove the offending label. Statistical sampling does, however, work very well in controlling the quality of printed material for poor printing, coloring, centering, etc.

Sampling and inspection of bottles, containers, and cartons has had wide application in the pharmaceutical industry. MIL-STD-105D works very well on this type of application since the vendor should understand the use of this plan.

It is not the statistical aspects of sampling plans which are of major concern in the above applications, but rather the definition of what is a defect. It is possible to have 20 or more types of defects for glass bottles. Are they all to have the same weight as to being called a defect? At what point is a defect called a minor defect, and when does it become a major or critical defect? If the inspection process is human, there is an ever-present tendency to tighter standards. Likewise, if a lot does not quite meet sampling plan standards, there is a tendency to go outside the sampling plan for information as to whether to accept or reject. If this item is needed for production due to low inventories or if the defects all run to 1 type which is not too serious there is a desire to accept the lot. Hence, the most important aspect of designing a sampling plan is a definition of what a defect is or what is to be sampled for. The most important aspect of using a sampling plan is drawing a truly random sample and following and maintaining the definition of a defect.

CONTROL CHARTS

Description of Control Charts.—Control charts are a method of graphing or plotting data in such a manner that a time series of events is summarized. Figure 4 shows a control chart described by Breunig.

Control charts are characterized by a vertical axis which has a scale of a varying measurement such as a mean, range, standard deviation, or fraction defective, and a horizontal axis which is time-oriented. There are frequently 3 horizontal lines on a control chart. The center line is the target value or the historical process average. The upper line is the upper control limit (UCL) which is normally 3 standard deviations above the center line. Likewise, the lower line is the lower control limit (LCL), again 3 standard deviations below the center line. The 6 standard deviation spread between the upper and lower control limits will encompass 99.97% of the values in a distribution with its mean at the

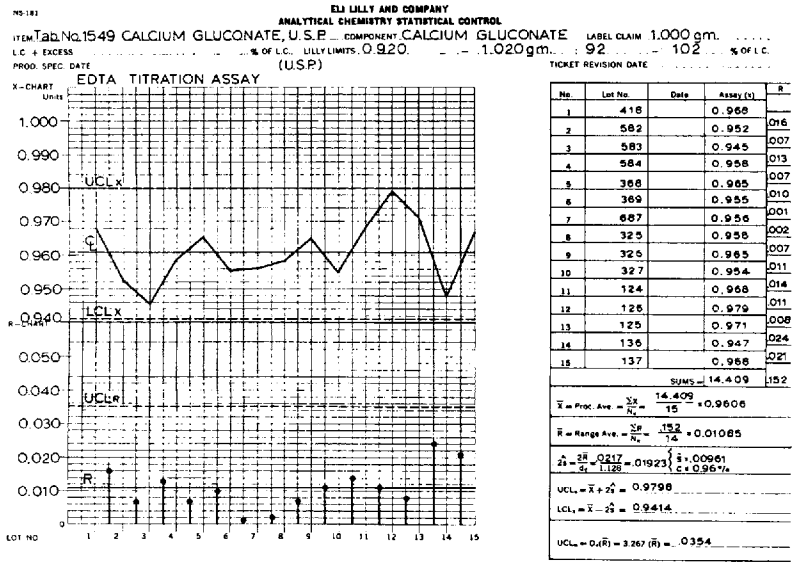

Fig. 4.—A typical analytical laboratory control chart. (Breunig.)

center line or process average. Control charts which are used for plotting sample averages are frequently used in conjunction with a range or standard deviation chart. One chart is used to plot the process average and the other the process variability. If a given value falls outside of the upper or lower control limit, the process is usually adjusted to bring it close to the center line.

It is frequently desirable in pharmaceutical processes to add another set of limits which could be called tolerance limits. If a given value falls outside the tolerance limits, the operation must be halted. If a tablet compressor normally operates under control limits well within the U.S.P. and N.F. specifications, the tolerance limit might well be the equivalent of the U.S.P. and N.F. specifications. This ensures the manufacturer that few, if any, products are manufactured outside of acceptable limits.

There have been many modifications to the structure of control charts, notably the cumulative sum type. However, the use of control limits for decision making is common to all types of them.

Application of Control Charts.—Control charts are primarily used for plotting routine (*a*) analytical or biological assay results or parameters, (*b*) fills of fluid or injectable products, (*c*) weights of tablets or capsules, and (*d*) the percentage or number of defects in a sample of packages emanating from a packaging operation. They are also useful for plotting data gathered at the beginning of a research problem.

Yehle (30) in an investigation of laboratory precision and specification limits used a type of control chart to plot the results of laboratory analyses.

Noel (31) discussed the use of control charts as applied to collecting data about particulate matter in ampuls in order to check the difference between machines and of fill weights. Noel's discussion on control charts was the second of a 5 part series of articles dealing with the use of statistics in the pharmaceutical industry (32–35).

Recently, Breunig (29) discussed the application of statistical control charts to analyses which were run in duplicate on vitamin A palmitate where the problem was *cis–trans* isomerization. Product and analytical control was obtained by purchasing an equilibrium mixture of *cis–trans* vitamin A palmitate. He also discussed the use of control charts for studying tablet weight variation and assay variation prior to introducing single tablet assays.

Brochmann-Hanssen and Medina (36) described the uses of control charts for determining weight variation and composition variation for phenobarbital tablets. Their inspection of the charts indicated that some disruption of uniformity occurred during the compression of the tablets.

One of the problems with control charts for routine applications is the objection of the analyst or operator to writing the results of a test on a control chart since the result is usually first written on another document. Wherever a control chart is a secondary document, resistance to its use is apt to appear. If the control chart can be the primary document, acceptance is made easier.

Breunig suggests as a solution to this problem

an instruction manual, a training course, consultation, and management support. Breunig also suggested that the reason analytical laboratories have not adopted statistical quality control programs is because the U.S.P. and N.F. have their sampling and control procedures "couched" in terms which can be considered obsolete in the light of the knowledge available today (29).

The above statement was in reference to the U.S.P. and N.F. attribute sampling plans for tablet and capsule weight and assay variation, the single assays performed on composite samples, and the associated unrealistic limits specified by the monograph.

EVOP AND ADAPTIVE QUALITY CONTROL: STATISTICS FOR CHEMICAL PROCESS EVALUATION

Quality control whose main forte is controlling new and existing operations, must at times be concerned with improving existing manufacturing operations. After publishing extensively (37–40) in the field of response surface statistics, Box (41) published a paper dealing with a very simple technique called "EVOP" (evolutionary operation) which was useful in finding the optimum operating conditions for a chemical batch process or similar process. His first paper dealt with 2 variables (*e.g.*, concentration and temperature) and 3 variables (*e.g.*, concentration, temperature, and pH) models. Without developing the theory of EVOP, a 2-variable EVOP cycle would consist of 5 experiments performed under the conditions shown in Fig. 5. By some very simple calculations (42), it can be established whether the process is at its optimum or whether the increase of one or both variables in a given direction tends to give a significant increase in response. Usually 3 cycles or 15 experimental units are required to show which variables are significantly influencing the process and in which direction they should be adjusted to maximize the response (yield, impurities, or cost of the process).

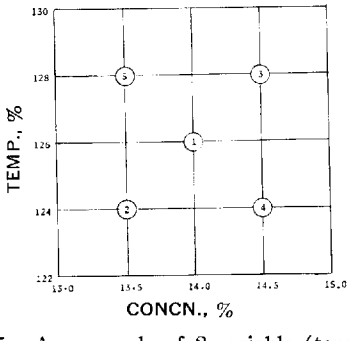

Fig. 5.—An example of 2-variable (temperature and concentration) EVOP cycle. [Figure 4 of Box (41).]

There have been other statistical models, such as adaptive quality control (42), developed for maximizing yield, but these are beyond the scope of this paper.

The point to be made about response surface statistics, as applied to making fine pharmaceutical chemicals or other material, is not with the statistical aspects of the technique, but with the control aspects.

EVOP and similar response surface techniques are difficult for the quality control function of a company to evaluate because they are opposed to the basic control precept of making the product by the same method under the same conditions for every batch. In conclusion, EVOP and related techniques are very useful and powerful statistical tools which definitely should be used on applicable production processes. Nonstatistical problems, however, may play a predominant or controlling part in the extent of the usefulness of these statistical production process optimizing techniques in the pharmaceutical industry.

BIOLOGICAL ASSAYS

Description of Biological Assay.—There are two types of indirect bioassay: one is based on quantitative responses, the other on quantal responses.

An example of a quantitative assay is the guinea pig skin test in which the activity of tuberculin P.P.D. is estimated from the varying size of wheals resulting from the injection of tuberculin-sensitized pigs with graded doses of the antigen.

An example of a quantal response bioassay is a mouse potency test in which antigenicity of a vaccine is measured from the proportion of mice still living after first being injected with the vaccine under test and then being challenged with the live organism against which the vaccine is to protect.

Biological assays can be further divided into two types: analytical dilution assays and comparative assays. In an analytical dilution assay (either quantitative or quantal), both a standard (or reference) and test preparation are considered to consist of an "effective constituent" suspended in an inert diluent (43). Relative potency in such an assay is defined as the ratio of the concentration in an effective dose in a unit amount of the test preparation to that of the standard. It then follows that the postulated relative potency characterizes the two preparations and

not the biological systems which are used to assay them (44).

Assays in which the postulated relative potency is not a constant are often referred to as "comparative" (43). A variable potency may occur in the comparison of a test and standard preparation which are qualitatively different. There are cases where the requirements of an analytical dilution assay are considered to be met and yet the outcome of the assay is far from being analytic because of the unknown extraneous factors which have caused qualitative differences in either preparation. (Analytical dilution assays are usually characterized by the reference and unknown having different slopes.)

In most biological assays, primary concern is directed at estimating the (*a*) effective dose of a test material in terms of the known effect of its constituents and (*b*) precision of that estimate.

Quantal Response Assay.—Finney (45) has pointed out that the ED_{50} (effective dose required to produce a 50% response) may have 1 standard error between determinations by 1 worker, another standard error between different workers using the same apparatus and the same stock of animals in a given laboratory, and still another between determinations in different laboratories. However extensive the experimentation on one population, no statistical analysis can demonstrate the applicability of the conclusion to a different population.

Finney further states that from the basic concept of plot technique, it can be deduced that the magnitude of an assay error is inversely proportional to the degree of correlation among the test subjects. An estimated internal assay error is necessarily an indication of what would be found in repeated tests of the same stock of subjects under the same conditions of testing. This limited interpretation must always be attached to statements about the ED_{50} of unknown material.

Nevertheless, much of the present difficulty seems to be removed by an adoption of comparative experiments (46). The numerous bioassay data which have been available to the writer appear to confirm Finney's statement that a theoretical reason exists for believing a numerical result to be applicable more widely than in the circumstances of the particular experiment that produced it.

The considerations that influence the interpretation of assay validity are of two types: statistical and other. Under statistical considerations, criteria such as linearity and parallelism of the dose–response relationships are included. The nonstatistical considerations include such matters as environmental factors which may have affected the quality of the effective constituent of either preparation. Obviously, both considerations are necessarily integral and eventually lead to a determination of whether the irregular occurrence is merely due to chance.

Such an extraneous factor as a physical disturbance apparently has an effect on the antigenicity of a certain type of vaccine to the extent that the effective constituent in the test material is no longer comparable to the standard. However, this type of deviation from parallelism does not necessarily imply that the assay is invalid (44).

From the realistic point of view, an assay may be considered to be invalid if and only if statistically significant nonlinearity and nonparallelism are accompanied with such factors as inconsistent testing conditions and heterogeneous test subjects which are not randomly allocated to each preparation.

In numerous cases of quantal response assays (6 points with approximately 16 subjects at each point), the test of linearity and parallelism often turns out to be so insensitive that the slope of the dose–response curve for 1 preparation must be at least 2.5 times steeper than that for the other preparation in order for the departure from parallelism to be statistically significant.

In such cases, neither the choice of the form of the transformation of the biological responses nor the arbitrary increase in the number of test subjects would overcome this difficulty. This type of insensitivity evidently stems from the inefficient allocation of dose levels which would be reflected by the insensitive responses.

It has been well-established that the relative potency, slope of the dose–response curve, and assay error represent characteristics of the test material (47, 48). In order for the subjects to be able to produce the true characteristics of the test material, it is a prerequisite for the experimenter to allocate an optimal amount of effective constituent in such a way that the minimum effective dose can be detected with a minimum of error.

Cramer (49) compared methods of fitting the dose–response curve for small samples (3-dose test with 10 subjects at each dose and 5-dose test with 4 subjects at each dose) and found that the probit analysis is superior to the minimum normit χ^2 method with respect to the assay error. This superiority is found to be substantial if the dose levels are poorly placed with respect to the true assay parameters.

The relative merits of the various methods for fitting the dose–response curve would be more meaningful if the existing antigenicity tests (mouse potency tests for influenza vaccine and

pertussis vaccine) are biologically unbiased. For instance, the following question may be asked: Is the antibody level at the time of challenge directly proportional (by an approximately constant factor) to the amount of a vaccine injected at the time of immunization?

It should be pointed out that many statistical analyses of quantal response bioassay are being utilized under the assumption that the reply to the above question is positive. This type of question further complicates the situations in which the valid evaluation of potencies of polyvalent vaccines are required.

Quantitative Bioassay.—The fundamental concepts involved in quantitative bioassay are essentially the same as those discussed under *Quantal Response Assay*. The quantitative assay in general, however, is subject to greater control and in turn requires more elaborate assay plans than the quantal response case.

Some of the antibiotic assays and vitamin assays consist of a number of plates, each plate being regarded as block. The different positions within a plate may have different effects on the response which should be solely attributed to a given concentration of the test material. Each sample should be rotated or allocated at random within a plate, depending on whether the position-to-position difference is systematic across the plates or unique to the individual plates. Both the test and the standard preparations should be represented in each plate, even in the cases where the position difference can be neglected.

It is well known that, when the dose and response are suitably transformed, the mode of the action of organisms can be represented by the monotonic dose–response relationship (either approximately sigmoidal or straight). Whenever a prior knowledge suggests that the dose-response relation for a given test material is approximately sigmoidal or linear, it is advisable to select two reference (standard) concentrations which cover the linear portion of the dose–response curve. The difference in response which is due to the difference in concentration is best detected along the linear portion of the dose–response curve.

The only remaining problem then is to choose target dilutions for the unknown preparation in such a way that at least 1 dilution for the unknown produces the response which stays within the responses produced by the 2 reference concentrations (50).

Another type of quantitative bioassay is that in which the response is measured by an instrument with a response in units of per cent transmission. Because of the possible instrumental drift which may be linear or nonlinear, the time required to process a set of standard and unknown dilutions may be regarded as a block.

There is still another type of quantitative assay in which the potency is measured from time. The activity of prothrombin samples, for example, is estimated from the clotting times for the standard and unknown samples.

Loomis (50) developed an assay method for fibrinolysin which uses a simple formula for estimating activity. His method is based on the assumption that the concentration–time relation for the unknown samples and standard is not only linear but also parallel to each other within the dilution range.

Stone and Bruce (51) reported (based on 250 humans injected intradermally with tuberculin) that there exists a linear relation between the induration–area and the log–dose in constant volume. However, they pointed out that the following questions deserve further clarification. (*a*) Can the form of the dose–response relation of intradermal tests (tuberculin) be extended to other diseases such as diphtheria toxin? (*b*) Is the form of the dose–response relation of intradermal test dependent on experimental technique?

They also presented a method by which the antibody level may be estimated from a knowledge of the dose and concentration of tuberculin used and the resultant area of induration.

The discussion under *Biological Assays* was included in this paper merely to show that it is a part of quality control. The subject alone is so large that a journal, *Biometrics*, is published to keep up with the changes in this one specialized field of statistics.

U.S.P. XVII (54) now recognizes the subject of biologics and biological assay and gives them prominent attention, as can be traced historically from 1950 to 1965.

U.S.P. XIV (20) had a single paragraph dealing with the significance of the standard error in a biological assay. There was no mention of the method of assay for antibiotics; antibiotic products were said to comply with the requirements of the Food and Drug Administration.

U.S.P. XV (52) included a section on the "Design and Analysis of Biological Assays," by C. I. Bliss. In addition, a section was added detailing the methodology for performing assays on the antibiotics included in the U.S.P.

U.S.P. XVI (53) expanded the section on antibiotic assays to show the formula necessary to calculate the slope and potency of the material being assayed. The section on the "Design and Analysis of Biological Assays" was continued.

U.S.P. XVII (54) again expanded the section on antibiotic assays to include some of the tech-

niques of designing antibiotic-type assays. Some revision was made to the section on "Design and Analysis of Biological Assays," including a glossary of symbols used within the section.

In a period of 15 years, the U.S.P. has gone from a book which included no statistical methodology dealing with biological products, to a book which from a legal point of view is an authority on how statistics are to be applied. Included in the section on "Design and Analysis of Biological Assays" is an official method for rejection of outlying or aberrant observations and the replacement of missing values. Likewise, the sections on the combination of independent assays and joint assay of several preparations are of great value to the industry. For the first time, a glossary of statistical symbols with official status is printed for all to use. The writer has a personal preference for a different set of symbols, but the benefits to be gained within the industry from a uniform set of symbols and formulas more than offset any personal feeling.

Beginning in U.S.P. XVI (53), a small section was added called "Biologics," which simply stated that for vaccines, antitoxins, etc., the National Institutes of Health, Division of Biological Standards, controlled the testing requirements and that the actual assay procedure was beyond the scope of the book.

Many methods are commonly used for computing the potency of biological products, some of which are defined adequately by the National Institutes of Health. However, there is a definite need for a specification of the statistical method of choice, regardless of the length of calculation, and very definite rules for the determination of assay validity, combination of independent assays, and joint assay of several preparations.

Bliss' section on the combination of independent assays states that: "additional animals can be added to an insufficiently precise assay until the combined results reduce the confidence interval within the lines specified in the monograph."

Since many biologics are assayed on the basis that the unknown is equal to or greater than the reference standard in potency, it is logical for a manufacturer to make the potency of the vaccine just potent enough to exceed the reference a certain percentage of the time. If the manufacturer performs 4 assays on the unknown and combines their results, which in reality doubles the precision of the assay, this then allows the manufacturer to produce a product whose potency tests greater than the reference about the same percentage of the time, but on a comparison basis does not have to be so excessively potent as material which is only controlled by a single assay.

If the Division of Biological Standards performs a single assay on material which has had multiple assays performed on it by a manufacturer, there is a high probability that the material will fail; hence, the lot is not suitable for distribution. If the Division of Biological Standards reference standard is to be a standard of potency to which a vaccine is to be equated, then the manufacturer should be allowed to use all available statistical methodology to warrant that the product is equal to or greater than the standard.

SUMMARY

Mathematical statistical methods have been found to be useful through all areas of quality control. Statistical sampling plans applied to dosage form variation, by their very controversial nature, have exposed large segments of the pharmaceutical industry to statistics. Analysis of variance, design of experiments, and hypothesis testing have not in themselves found extensive usage in quality control functions because of their research orientedness. Applications using these methods apply and are very effective when dealing with materials which are not routinely meeting their specifications or where specifications are in the process of being developed. Biological assays are now having increased attention paid to their problems, as witnessed by the U.S.P. sections dealing with biologicals.

Most statistical techniques are based on a measure of accuracy (the mean of a set of data) and a measure of variation (the standard deviation). With the advent of the food and drug law amendments of 1962, the increased emphasis on good manufacturing practice, the introduction of high-speed packaging lines, dosage form variation, glass bottle defects, missing bottle labels, and granulation characteristic variation, etc., the concept of statistical variation is no longer welcome to the quality control function. Statistical methodology says there will always be defects. Quality control says there must be zero defects.

The next 15 years will be interesting to see how well the statisticians can adapt themselves to problems implying sample sizes in the 500 to 10,000 range rather than 1 to 1,500.

The statistician and the quality control man must bear in mind the truism, "To attain quality, one must first measure it, and to measure quality, one must establish a standard of rejection." As Youden has pointed out, statistics are the laws of measurement. The new era for quality control statistics may well be in product design, control system design, or quality control simulation—all things to be done before the product is ever manufactured.

REFERENCES

(1) Feldmann, E. G., *J. Pharm Sci.*, **52**, I, November (1963).
(2) Mainland, D., New York University Medical Center Notes No. 33 and 34, May 1, 1963.
(3) Youden, W. J., *J. Assoc. Offic. Agr. Chemists*, **45**, 169 (1962).
(4) Garrett, E. R., *J. Pharm. Sci.*, **51**, 672(1962).
(5) *Ibid.*, **51**, 764(1962).
(6) *Ibid.*, **51**, 767(1962).
(7) *Ibid.*, **51**, 1034(1962).
(8) *Ibid.*, **51**, 1036(1962).
(9) Davies, O. L., *Technometrics*, **1**, 49(1959).
(10) Calder, A. B., *Appl. Stat.*, **9**, 170(1960).
(11) Wernimont, G., Presentation at P.M.A. Research and Development Section, November 1963.
(12) Breunig, H. L., and King, E. P., *J. Pharm Sci.*, **51**, 1187(1962).
(13) Wald, A., "Sequential Analysis," John Wiley & Sons, Inc., New York, N. Y., 1947.
(14) "Military Standard Sampling Procedure and Tables for Inspection by Variables for Percent Deviation (MIL-STD-105D)," Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C., April 1963.
(15) "Military Standard Sampling Procedure and Tables for Inspection by Variables for Percent Deviation (MIL-STD-414)," *ibid.*
(16) Duncan, A. J., *Assoc. Am. Fertilizer Control Officials, Offic. Publ.*, **13**, 42(1959).
(17) Duncan, A. J., *J. Assoc. Offic. Agr. Chemists*, **43**, 831 (1959).
(18) Duncan, A. J., *Technometrics*, **4**, 319(1962).
(19) Duncan, A. J., "Operating Characteristics of Fertilizer Inspection Plans Based on the Miles-Quackenbush Tolerances Illustrated for Mixed Fertilizer with 10% Nitrogen," 1959 Proceedings of National Plant Food Institute Conference.
(20) "United States Pharmacopeia," 14th rev., Mack Publishing Co., 1950, pp. 689, 799; "National Formulary," 9th ed., Mack Publishing Co., Easton, Pa., 1950, p. 787.
(21) Dunnett, C. W., and Crisafio, R., *J. Pharm. Pharmacol.*, **7**, 314(1955).
(22) Smith, L. K., *ibid.*, **7**, 875(1955).
(23) Green, M. W., and Knudsen, L. F., *J. Am. Pharm. Assoc., Sci. Ed.*, **39**, 599(1950).
(24) Haynes, J. D., and Schnall, M., "Dry Sterile Solids Weight Variation and Recommendation for Specification," March 27, 1963, unpublished report.
(25) Paul, J. F., to P.M.A., Quality Control Section subcommittee chairman, unpublished report.
(26) Moskalyk, R. E., Chatten, L. G., Cox, C. E., and Pernarowski, M., *J. Pharm Sci.*, **50**, 651(1961).
(27) Haynes, J. D., Schnall, M., and Lamm, R. A., "Variability Tests in Acceptance Sampling from Non-Normal Populations," presented to the American Statistical Association meeting, Cleveland, Ohio, September 1963.
(28) Article 5, Department of Agriculture, State of California, 1961.
(29) Breunig, H. L., *Industrial Quality Control*, August, 79(1964).
(30) Yehle, E. C., *Anal. Chem.*, **25**, 1047(1953).
(31) Noel, R. H., *Drug Allied Ind.*, **38** (April), 19(1952).
(32) *Ibid.*, **38** (March), 14(1952).
(33) *Ibid.*, **38** (May), 15(1952).
(34) *Ibid.*, **38** (July), 21 (1952).
(35) *Ibid.*, **39** (February), 14(1953).
(36) Brochmann-Hanssen, E., and Medina, J. C., *J. Pharm. Sci.*, **52**, 630(1963).
(37) Box, G. E. P., and Wilson, K. B., *J. Roy. Stat. Soc. (b)*, **13**, 1(1951).
(38) Box, G. E. P., *Biometrics*, **10**, 16(1954).
(39) Box, G. E. P., Connor, L. R., Cousins, W. R., Davies, O. L., ed., Himsworth, F. R., and Sillito, G.P., "The Design and Analysis of Industrial Experiments," Oliver and Boyd, Edinburgh and London, 1954.
(40) Box, G. E. P., and Youle, P. V., *Biometrics*, **11**, 287 (1955).
(41) Box, G. E. P., *Appl. Stat.*, **6**, 3(1957).
(42) Box, G. E. P., and Hunter, J. S., *Technometrics*, **1**, 77(1959).
(43) Burn, J. H., Finney, D. J., and Goodwin, L. G., "Biological Standardization," 2nd ed., Oxford University Press, London, England, 1950.
(44) Cornfield, J., *J. Pharmacol. Exptl. Therap.*, **144**, 143 (1964).
(45) Finney, D. J., "Statistical Methods in Biological Assay," 2nd ed., Hafner Publishing Co., New York, N. Y., 1964.
(46) Finney, D. J., "Probit Analysis," 2nd ed., Cambridge University Press, London, England, 1962.
(47) Cornfield, J., and Mantel, N., *J. Am. Stat. Assoc.*, **45**, 181(1950).
(48) Bliss, C. I., *J. Pharm. Pharmacol.*, **11**, 192(1938).
(49) Cramer, E. M., *J. Am. Stat. Assoc.*, **59**, 779(1964).
(50) Loomis, E. G., George, C., Jr., and Ryder, A., *Arch. Biochem.*, **12**, 1(1947).
(51) Stone, M., and Bruce, R. A., *Biometrics*, **17**, 33 (1961).
(52) "United States Pharmacopeia," 15th rev., Mack Publishing Co., Easton, Pa., 1955, pp. 848, 865.
(53) *Ibid.*, 16th rev., 1960, pp. 814, 859, 873.
(54) *Ibid.*, 17th rev., 1965, pp. 768, 832, 843, 905, 919.

Research Articles

Effects of Potential Inhibitors on Metabolism of Griseofulvin *In Vitro*

By S. A. KAPLAN, S. RIEGELMAN, and K. H. LEE

The inhibitory effects of *p*-ethoxyacetanilide, *p*-methoxybenzylamine, codeine, and SK&F 525-A on the metabolism of griseofulvin were studied in a Krebs-Ringer bicarbonate liver-slice system. This investigation sought to determine the possibility of prolonging the biological activity of griseofulvin.

GRISEOFULVIN, the first available oral antifungal antibiotic, was originally isolated from the mycelium of *Penicillium griseofulvum* by Oxford in 1939 (1). However, it was not until 1958 that Gentles (2) reported that he was able to eradicate experimental ringworm of guinea pigs as a result of oral treatment with griseofulvin. This was followed by the works of Riehl (3, 4), Blank *et al.* (5), and Williams (6), who obtained favorable results in the treatment of superficial fungi infections of man.

Received August 13, 1965, from the School of Pharmacy, University of California, San Francisco Medical Center, San Francisco.
Accepted for publication September 15, 1965.
Abstracted in part from a thesis submitted by Stanley A. Kaplan to the Graduate Division, University of California, San Francisco Medical Center, in partial fulfillment of Doctor of Philosophy degree requirements.
This investigation was supported in part by grants AI 05241 and 5 TI GM 728 from the U. S. Public Health Service, Bethesda, Md.
The authors are grateful to Miss Darlene Tanneberg for technical assistance.