NBSIR 75–697

CONSIDERATIONS IN THE USE OF SAMPLING PLANS FOR EFFECTING COMPLIANCE WITH MANDATORY SAFETY STANDARDS

V. L. Broussalian
A. J. Farrar
J. W. Lyons
C. O. Muehlhause, Chairman
M. G. Natrella
J. R. Rosenblatt
R. D. Stiehler
J. H. Winger

June 1975

U.S. DEPARTMENT OF COMMERCE, Rogers C.B. Morton, Secretary

NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, Director

CONSIDERATIONS IN THE USE OF SAMPLING PLANS FOR EFFECTING COMPLIANCE WITH MANDATORY SAFETY STANDARDS

June 1, 1975

NBS Committee:

V. L. Broussalian
--Technical Analysis Division
(Currently: Federal Energy Agency)

A. J. Farrar
--Legal Advisor

J. W. Lyons
--Center for Fire Research

C. O. Muehlhause, Chairman
--Institute for Applied Technology

M. G. Natrella
--Statistical Engineering Laboratory

J. R. Rosenblatt
--Statistical Engineering Laboratory

R. D. Stiehler
--Center for Consumer Product Technology

J. H. Winger
--Center for Fire Research

CONTENTS

CONSIDERATIONS IN THE USE OF SAMPLING PLANS FOR EFFECTING COMPLIANCE WITH MANDATORY SAFETY STANDARDS

I. Introduction

Though a consumer may be concerned about his own or his family's safety in using contemporary products, he has neither the knowledge nor the resources to perform an adequate safety evaluation every time he purchases a product. A growing body of Federal regulations is aimed to insure the public that safety--especially safety from invisible hazards--is guaranteed by the establishment and enforcement of mandatory standards.

In commerce regulated by mandatory safety standards, the manufacturer wants to minimize the cost of compliance and the consumer wants to be sure that standards have been met, while both want to minimize the expenditure of tax dollars to supervise the operations of industry. Thus all parties can benefit from the wise use of compliance procedures engaged in by manufacturers and regulatory officials. In recent months, the Consumer Product Safety Commision (CPSC) has been re-examining the role and the application of sampling plans to mandatory safety standards for purposes of effecting compliance.*

A performance standard begins with the choice of a method for measuring (and assigning a numerical value to) the quality of a product, and then specifies the range of values within which measurements should lie for conforming products. A "sampling plan" is a statistical prescription that specifies procedures for judging the quality of a group of products on the basis of measurements made on just a few of the items in the group. For example, in certain simple situations the plan may state how many samples from a given size production lot must be tested and, of these, how many must "pass the test, i.e., standard" for the lot to be found acceptable. The sampling of products for purposes of determining their acceptability can be performed by the producer before release of his products to the public, by the regulator from products selected in the market place from "off-the-shelf," or by some combination of both these sampling methods.

---

*"Hearings before the Sub-Committee on Commerce and Finance of the Committee on Interstate and Foreign Commerce, House of Representatives: Reviewing the Operations of the Consumer Product Safety Commission since its Inception, May 14, 1973." U.S. Government Printing Office, Washington, D.C., 1974.

A number of Federal agencies are experimenting with the use of sampling to assure or oversee compliance with product safety standards in areas not traditionally regulated. These include the Department of Transportation, the Labor Department (Occupational Safety and Health Administration), and the Environmental Protection Administration.* Many of the examples in this report refer to the CPSC, however, because of the NBS experience with flammable fabric standards, now administered by the CPSC.

Under the Flammable Fabrics Act (PL 90-189) a precedent was set by the Department of Commerce to include sampling plans in standards for mattresses and for children's sleepwear. In spite of this, the authority to mandate a plan as part of a standard for other products falling within the scope of the Consumer Product Safety Act has been questioned by the Commission's Oversight Committee, the House Sub-Committee on Commerce and Finance of the House Interstate and Foreign Commerce Committee. In addition, various industry spokesmen as well as various "consumer advocates" have challenged the wisdom of mandating a specific plan rather than following some other means of assuring compliance. This is not to say either all industry or all consumer interests oppose (or favor) sampling plans. These groups are, in fact, split on the issue, as may be seen by examining the record of a public hearing held on this matter by the Commission April 4-5, 1974.

To develop some insight into this general problem an NBS-wide Committee was formed composed of staff having technical, statistical, economic, and legal backgrounds, and having also some knowledge of the means by which standards are generated. The committee was directed to formulate and discuss the role of sampling plans in compliance testing, and, in particular, to consider the relationship of sampling plans to mandatory product safety standards. This report is a result of that committee's effort.

Following a general exposition in Chapter III, the last two chapters view the subject matter from special vantage points. Chapter IV on cost/benefit analysis constitutes an attempt to synthesize the principal factors affecting the public interest into a comprehensive view. Chapter V on the use of sampling plans in compliance testing, examines the technical questions and policy issues that a regulatory agency must resolve in order to develop a suitable set of statistical sampling procedures in any particular instance.

---

*"ASQC and the National Standards Scene-Harvesting the Mandatory Mushrooms," Robert A. Abbott, Technical Director ASQC, Presentation to the American Society of Quality Control, October, 1974.

II. Summary and General Conclusion

1. After identifying a hazard which he deems is in the public interest to reduce, the regulator begins by devising a technique for measuring some appropriate physical feature of the product which is suspected of causing injury. For example, in the case of fabrics, he may determine that the length of a char induced by a suitable ignition source correlates well with flammability and the risk of burn injury. Having made such a determination, the regulator is then in a position to specify an objective means of measurement and the setting of a proper standard. The latter consists of specifying a range of acceptable values of the product feature, e.g., flammability when it is measured as prescribed by the standard. It must be appreciated that any product, even though it meets the standard, still contains some residual level of risk.

2. Having established a "reasonable" standard by taking into account costs, benefits, and the residual risks which people are willing to assume, the regulator must further find an effective means of gaining compliance. The five which follow constitute his major available options. In the order of taking earlier and earlier action in the production process for the purpose of avoiding injury from unwanted products they are the following: (1) processing of complaints about injurious and defective products leading; for example, to recall, to other administrative action, or to support of liability litigation; (11) off-shelf market place sampling of products leading to further action such as an investigation of the manufacturer's quality control procedures or to direct administrative legal action; (III) voluntary sampling or other quality control assurance provided by the producer; (IV) in-plant mandatory sampling prescribed by the regulator; and (V) prototype testing prior to production. Combinations of these can also be used.

3. Methods (I) and (II) are best adapted to the detection of gross non-compliance which would result from fraud, ignorance, or incompetence on the part of the producer. They have little capability to detect marginal or intermittent non-compliance. A principal reliance on these methods has the disadvantage of invoking costly and difficult remedies for non-compliance, including, for example, recall and the threat of criminal proceedings against the producer. Nevertheless, these methods of checking compliance must remain available to the regulator. They also serve as a basis for examining the control procedures utilized in the manufacturing plant (method III).

4. Methods (III) and (IV) constitute the taking of action to head off non-compliance and possible injury at an earlier stage of the production process. Costs are not avoided by these methods, but rather shifted from recall to more extensive testing, additional quality control, and the rejection or reworking of production lots containing defective products which fail to pass the standard as measured on a production line.

5. Though method (IV) constitutes considerable intervention and control by the regulator, it is nevertheless capable of gaining two major objectives:

(1) statistically reliable evidence for the regulator that a reasonable fraction of products released to the market place meet the standards; and

(2) an incentive for the producer to observe his end product quality, and from this to take prompt, appropriate action to avoid costly rejects in achieving the desired quality.

6. Method (V) constitutes the earliest intervention in the production process by the regulator. In this method prototype production models are examined by the regulator for satisfactory safety performance. If found to be adequate they are approved for production.

7. In effecting compliance through statistical sampling procedures, the regulator recognizes that no sampling scheme, even 100 percent testing, can eliminate all products which do not meet the standard. Just as zero residual hazards are nearly impossible to achieve in any product, so is zero defect production, as defined by the standard. Residual hazard and residual defects are both inherent characteristics which must be accounted for in any control scheme which might be utilized.

8. That a reasonably small fraction of non-comforming products will reach the market place, even under strict compliance with a prescribed sampling plan (method IV) does not necessarily mean that a producer will be exempted from civil liability for damages which his product may have caused. On the other hand, the producer might reasonably defend himself against criminal liability charges or other administrative action brought by the regulator if he could demonstrate full compliance with the sampling plan and substantial compliance with the standard (i.e., few products not meeting the standard and finding their way to the market place). The matter, however, of civil and criminal liability regarding those products that do not meet the standard are, of course, substantive legal questions which are beyond the scope of this report and are for resolution and disposition by the courts and the regulatory system.

9. From a cost/benefit point of view, an analysis shows how a hypothetical regulator, knowledgeable of both production and damage costs, could set a standard and sampling plan in such a way as to optimize the net public benefit. In performing this task, the regulator must take into account the fact that without the standard the producer is not always held liable or accountable for all of the damages which his product causes. In addition, the regulator must reckon properly with the uncertainties present in both the production process and in the performance measurements required to test the product for acceptability. In complying with the standard and sampling plan provisions, the producer is motivated to achieve acceptable quality by avoiding costly rejections of his end product. Cost/benefit analysis shows precisely how this could automatically come about.

10. From a statistical point of view a discussion of various approaches to the design of sampling plans shows how sampling methods may be tailored to the special circumstances of a particular application. Statisticians stress the fundamental distinction between product requirements (defining conformance) and a sampling plan (for judging conformance). Sampling inspection practice is based on the use of sampling "schemes"--sets of sampling plans put together into systems with accompanying procedures characterized by some desired overall objective in the way of quality assurance. Existing sampling schemes were devised for large-scale procurement, primarily military, and are not appropriate for compliance testing, even though the sampling plans used as building blocks in those acceptance sampling schemes can also be used in schemes designed for compliance testing. Until a variety of general schemes has been demonstrated for utilizing sampling plans in compliance testing, sampling procedures must be designed on a case-by-case basis.

11. In practice the regulator may employ more than one of the alternatives, allocating resources among them in the most effective way.

The advantages and disadvantages of the various alternatives can be summarized as follows.

Regulator takes responsibility for testing (Alternative I and II)

Advantages:

° Allows limited resources of regulator to be concentrated on products from incompetent or unethical manufacturers.

° Is least costly for products where design predominantly governs safety.

° Interferes least with manufacturing operations, but places responsibility on the manufacturer.

° Penalizes violator of law without penalizing good manufacturers.

Disadvantages:

° Is not suitable for products where process control is critical for safety, such as chlorination of drinking water or canning of foods.

° Is not effective for detecting variable product quality that affects safety; for example, flame retardent treatment of textiles.

° Acts after-the-fact and is not preventive inspection for defective products.

Manufacturer takes responsibility for testing (Alternative III and IV).

Advantages:

° Is major effective means for controlling safety of products where process control is critical.

° Detects and removes unsafe products before they reach the market place.

° Is most effective for obtaining uniform product quality.

° Shifts burden of testing from the regulator; requires less of his resources.

Disadvantages:

° Interferes most with manufacturing operations, particularly in the case of mandatory sampling plans.

° Can cause needless testing of products that comply with the standard.

° Increases costs to good manufacturers as well as to violators.

° Distributes limited resources of regulator over both safe and unsafe production.

Manufacturer develops/tests improved design (Alternative V)

Advantages:

° Is a highly preventive method.

° Is most effective in detecting inherent hazards in materials or systems.

Disadvantages:

° Is not suitable for controlling safety of products where process control is critical.

° Delays introduction of new products to the market place.

12. It is concluded that there are advantages and disadvantages to all of the options cited and that any one or a combination may be appropriate in a given instance. In particular, the regulator should consider the use of sampling both before and after a product is issued to the market place. In this way he may choose the most cost-effective method.

III. General Exposition

A. Setting the Standard. The Consumer Product Safety Commission and some other regulatory agencies are charged with the responsibility and authority for promulgating safety standards when it has been ascertained by them that it would be in the public interest to do so. For example, the Commission is directed to evaluate the severity and frequency of injuries which may result from a suspect product hazard and also the reasonableness of expecting that the consumer, if he fully understood the hazard, would accept the risks involved. In addition, the regulator is usually expected to make an evaluation of the economic consequences of any standard which he would propose.* The purpose of this analysis or evaluation is not simply to give due consideration to the producer per se, but also to give adequate weight to the utility or benefit which the public derives from the product or service. Generally though not always, as safety is improved the cost of production increases and prices rise accordingly. This has the effect of reducing or even curtailing the flow of products or services to the market place. It is therefore important for the regulator to be knowledgeable of the state-of-the-art of production so that the cost of anticipated changes to improve safety can be estimated. Economists sometimes employ a formal methodology termed "cost/benefit analysis" for balancing all of the factors cited. This approach is utilized in Chapter IV of this report; in Chapter III a more qualitative account is given.

Having examined the evidence provided by epidemiological studies, human behavior analysis, product engineering analysis, the etiology of accidents, and other related matters as indicated above, the regulator may conclude that the quality of safety can and should be improved in the product or service which he regulates. As an example, one may cite the case for reducing burn injury from children's sleepwear. In order to proceed, the regulator must first identify the product feature which needs to be modified if the hazard in question is to be reduced. In the case of fabrics, it was judged to be flammability.

Having identified the hazard with its associated product features, two principal courses of specifying a standard are in general open to the regulator. The first which we shall mention briefly in passing is the one of specifying a minimal standard for the product that relates primarily to design considerations. The standard in effect constitutes a prescription which the producer must follow in the design of his product. Evidence that he has followed this prescription is all that is needed for the producer to be found in compliance.

---

*Sometimes the Congress indicates that the cost of compliance is not to be considered in the setting of standards; however, in the case of the CPSC the costs are to be considered.

Such standards can be used when a) freezing the technology is not important in relation to the hazard to be avoided; b) the performance can be confidently anticipated from the design prescription; or c) the design is proven in advance of production by satisfactory demonstration of the prototype. Under these conditions, inspection for compliance is straightforward.

Consequently it is appropriate to shift consideration to the second course of action open to the regulator, namely the one of specifying a standard that relates primarily to performance. Such a standard can accomodate various designs and technologies by specifying a minimal performance expected of the product. Before such a standard can be specified, however, it is first necessary to translate the potential hazard into a physical measurement on the product. Thus in the case of flammable fabrics, a measurement of the length of a char induced by a suitable ignition source is utilized as a practical and proper means for measuring flammability. The validity of such an identification rests on having established a proper correlation between the characteristic measurement and the actual hazard in the field. In general, the identification of a useful means of measuring potential hazard or product quality constitutes a difficult problem. The National Bureau of Standards is frequently involved in this phase of the standards setting effort.

Having determined the means for measuring performance, i.e., with respect to the quality of safety at issue, the regulator can set the "standard" in terms of acceptable measurement values. Often this appears in a form which states that unless the quality parameter measured is less (or greater) than some specified value, the product is not acceptable.

For example, in the case of children's sleepwear, flammability is measured by "char length," the length of the burned area of a garment ignited in a specified way.* An average char length exceeding 17.8 cm (7.0 in.) is stated to be unacceptable. The choice of the numerical value which separates "acceptable" from "unacceptable" performance is made by a regulatory agency on the basis of an appropriate combination of technical, economic, and political considerations. These might include evaluations of the relationship between the performance measure and the risk of injury, of the economic feasibility of achieving "acceptable" performance at reasonable cost, and of consumer preferences. Weighing all such factors, the regulatory agency adopts one or more numerical values** that, along with the measurement method, completes the definition of acceptable products.

---

*"Standards for the Flammability of Children's Sleepwear", DoC FF 3-71, as amended, Federal Register Vol. 37, No. 141, July 21, 1972, pp. 14624-32.

**For example, again in the case of children's sleepwear, very few char lengths observed in a series of measurements on sample products from a production lot may exceed 25.4 cm (10.0 in.). (See also Appendix B.)

One of the technical considerations that arises in setting a standard is that the quality of production items is variable.

Some manufacturers may be more successful than others in reducing this variability in their products. Appendix A reproduces a discussion by Bartky of how product variability can be considered in choosing the definition of acceptable products and how in turn the setting of a standard may influence manufacturers to reduce product variability.

If the tests for quality were cheap and non-destructive, perhaps all products would be tested. Often, however, this is either not possible (as when destructive tests are required) or ill advised (as when testing is costly). Therefore in many cases the manufacturer of a regulated product will perform the prescribed tests only on a sample of production items. The regulator, in setting a standard, is obliged to consider to some extent how compliance testing can be performed by the manufacturer, and in particular how the latter might employ sampling techniques. In some cases, therefore, the regulator also considers regulating the use of sampling and inspection methods.

The intent of the standard can not be to eliminate all hazard or all uncertainty concerning the residual hazard which remains, as neither of these "objectives" are physically achievable in practical terms. Risk and uncertainty can only be reduced, not eliminated, and this only at a price; moreover, at some point the increased price becomes unacceptable in comparison to the reduction in risk. It is this point which the regulator, acting as surrogate for the public, is expected to determine. A formal treatment of these concepts concerning hypothetical regulators acting strictly according to cost/benefit methodology is given in Section IV of this paper, which treatment constitutes a novel effort to analyze explicitly the interaction between the technical and economic aspects of standards-making.

The regulator ideally seeks to set the standard and effect compliance in such a way as to achieve an optimum flow of benefits for the "least cost" and "least damage" to the personal health and safety of the public. When compliance testing is done by sampling, a few non-complying items will reach the market; however with careful quality control and sampling inspection there should be only rare and small departures from the quality level specified in the standard,

It must be appreciated that variations in the quality of manufactured items necessarily will occur. Thus, even though extensive testing of the quality of randomly chosen samples from a production lot may be made, no guarantee can be made nor expected that all of the items in the lot are of the quality one might wish. This would be true even under 100 percent testing.

When production is monitored consistently by a well-defined sampling scheme, it is possible to calculate the chance that an occasional substandard product will reach the market. The sampling procedures are designed to make this risk quite small. Additionally, the sampling plan gives the producer an incentive to improve quality and further reduce his own risk that production lots will have to be rejected. Evidence accumulated from repeated uses of a sampling plan will show that the quality actually achieved is usually better than that indicated by the risk level set in the design of the sampling plan. Section V of this paper discusses in greater detail the design and use of various sampling plans and inspection schemes.

Regulation of the use of sampling inspection may reinforce the economic incentive to comply with a standard, but regulations requiring specific sampling procedures as evidence of compliance with a standard impose costs on a manufacturer; the more testing, the greater the cost.

B. Alternative Means of Gaining Compliance. There are various methods which the regulator may utilize to gain compliance with his standard. Not mamu pf them employ mandatory sampling plans. At least five major alternatives are open to him, and these will be discussed below. They are taken up in order of earlier and earlier involvement by the regulator in the means employed by the manufacturer to effect compliance. In all cases it is assumed that the regulator can set a "standard" and a specified method of measuring acceptability of any one item.

Alternative I

THE REGULATOR RECEIVES AND PROCESSES COMPLAINTS REGARDING BOTH INJURY AND DEFECTIVE PRODUCTS. HE THUS PROVIDES AN ADDITIONAL MEANS FOR HOLDING MANUFACTURERS LIABLE FOR DAMAGES.

The regulator sets a mandatory standard and is in a position of authority to test and evaluate products claimed to be defective and/or to be involved in injury. The standard, resulting from administrative procedures under the applicable law or laws, plays a role in any legal contest about a manufacturer's civil or criminal liability. This mode of action is various, complex, and tailored to each case; however, it has the overall effect of bringing the accountability for damages home to the manufacturer a greater fraction of the time than would prevail in the absence of the standard. To some extent then, the previously unaccounted for damages have been made accountable or brought within the scope of responsibility* of the manufacturer. This in turn acts as an incentive on the part of the manufacturer to comply with the standard and thereby reduce the risk to the public.

---

*Economists term such a process: "internalizing an externality" (i.e., of damages formerly outside the system).

Alternative II

THE REGULATOR SAMPLES PRODUCTS FROM "OFF-THE-SHELF" IN THE MARKET PLACE AND DISCOURAGES PRODUCTION AND DISTRIBUTION OF NON-CONFORMING PRODUCTS BY THE IMPLICIT THREAT OF STRONG ACTION AGAINST THE MANUFACTURER OF PRODUCTS THAT MIGHT PRESENT DANGERS TO THE PUBLIC.

Processing complaints of injury and of defective products (Alternative I) is a procedure limited to after-the-fact conditions. If it were possible to avoid more of these unfortunate events, it would be well for the regulator to do so. Accordingly, the next step he can take in the direction of assuring compliance is to examine products on the market shelf prior to purchase and encounter by the consumer. With his technique for measuring acceptability, the regulator may spot check products, particularly those which may be suspect. Non-conforming brands or lots may be recalled and/or other actions taken within the framework of administrative law.

Though this action may add to the arsenal of compliance techniques which the regulator can exercise, it still has many limitations with respect to proper checking or sampling of all production lines. It is at this point that the regulator begins to face the "statistical facts of life," the principal one being that some fraction of non-conforming products (i.e., ones which would fail to pass the standard test) will reach the market place; moreover, he must regard that this may well not be due to any malevolence on the part of the producer, but rather be the result of uncertainties in measurement, production, and quality control. Earlier it was recognized that conforming or not, essentially all products retain some residual hazard, i.e., capacity for damage. Now the regulator must further reckon with the fact that not all non-conforming products (as defined by the standard) can be screened from the market place.

There are situations in which the measurements specified in a standard may be sufficiently easy and inexpensive (perhaps also non-destructive) to perform that it becomes a practical matter for the regulator to "spot check" for compliance by testing samples of products randomly taken from the market place. With his limited resources, however, the regulator cannot sample all of the production lots as might be done in-plant by the producer (Alternative IV). On the other hand, the cost of rejecting products after issue to the market place is greater than their rejection in-plant. Since a prudent manufacturer would try to avoid the risk of being found out of compliance by the regulator's inspector, market place testing is suitable primarily for the detection of gross non-compliance arising from fraud, ignorance, or incompetence.

Minor departures from strict compliance with a standard are difficult to detect by market place sampling. The regulator usually has only a small inspection force, and must sometimes decide not to take action in marginal

cases. This situation can provide an economic advantage to a manufacturer who finds he can profit by willfully committing small violations, and who is willing to take this unfair advantage.

From the manufacturer's point of view, regulation by market place sampling can appear to be arbitrary, unfair, and capricious. Even if a regulatory agency conducts sampling inspections in the market place in accord with specified rules, actions taken against violators who are discovered only occasionally (because of limited inspection resources) may in the short run penalize only some violators, and these perhaps heavily via the recall of many lots.

Spot checking and follow-up market sampling are essential tools for the regulator in investigating reported injuries or complaints. This may or may not be the principal enforcement tool, but must always be an option for his investigation of compliance.

Alternative III

THE REGULATOR RELIES UPON THE PRODUCER TO ESTABLISH QUALITY CONTROL METHODS ADEQUATE TO YIELD PRODUCTS THAT COMPLY WITH THE STANDARD. THE PRODUCER MAY OR MAY NOT PERFORM SAMPLING INSPECTION OF END ITEMS. THE REGULATOR MAY OR MAY NOT REQUIRE EVIDENCE THAT END-ITEM TESTING IS UNNECESSARY.

Moving still farther in the direction of heading off difficulties before they arise, the regulator may strive for quality assurance before products reach the market place. Thus, in addition to, or in lieu of market place testing, he may seek assurance from the manufacturer by examining the latter's quality control procedures. This requires that the regulator be particularly knowledgeable of the production process. Perhaps some spot checking is also performed on a few of the end products, at least by the producer who needs to do a certain amount of this anyway in order to be sure that he is in adequate control of production.

There exist a number of situations in which, through a combination of proper materials, proper design, and proper control over the techniques of production, one can be reasonably assured that the end products will exhibit the sought-for quality; i.e., that they would pass a standard test of acceptability. Control charts, for example, can be maintained at several stages in the production process to provide a sensitive indication of stability or trend in the final product quality.

In addition to the above, the producer may also find it necessary to monitor final product quality by end product sampling using procedures of his own design. By this, it is meant that the frequency of testing and the number of allowable non-conforming items among products sampled are of the producer's design, but that the test itself or the method of measuring individual items would be the standard one specified by the regulator.

A certain amount of such testing for compliance must be made in any case by the producer, at least in the initial phase of his business. Through such testing, together with quality control records, he may learn for himself what process variables require tighter control in order to yield conforming products. Once, however, the production process is under adequate control, the frequency of end product testing may be relaxed. Should production quality remain uncertain, as for example when the state-of-the-art is marginal, end product testing may need to be continued on a regular basis, even though it might not be one precisely formulated by the regulator.

Alternative IV

THE PRODUCER IS REQUIRED TO TEST HIS END PRODUCT ACCORDING TO A SAMPLING PRESCRIPTION OR PLAN SPECIFIED BY THE REGULATOR.

If finally, the regulator finds it desirable, feasible, and cost effective to gain still tighter control and assurance over the products produced by the manufacturer, he may impose a system of sampling and testing of these end items. In short, he may mandate a sampling plan along with the standard. This constitutes the principal method at issue in this report, i.e., whether it can stand as a proper and viable regulatory method for assuring compliance, at least under appropriate conditions.

The regulator may prescribe in-plant sampling in several different ways: (1) He may do it himself. For example, under the Wholesome Meat Act of 1967, the Department of Agriculture places inspectors in meat packing plants to conduct or supervise the performance of tests. (2) The regulator may require manufacturers to establish specified sampling procedures in order to qualify the labeling of products as conforming to a standard. (3) The regulator may incorporate a sampling procedure in his standard. (4) The regulator may require manufacturers to have samples of their products tested by an independent (third-party) testing laboratory. In all cases, the cost of testing (at least some of it) and the cost of rejected lots or non-conforming products fall immediately on the manufacturer.

Variants of this alternative have been seen to be quite different in their legal consequences, and hence with respect to costs that may be incurred in case of legal actions. (For further discussion, see page 16.)

The intention of a mandatory sampling plan is to provide an incentive for the producer to manufacture conforming products, by requiring destruction or reworking of all items in a production lot whenever the sample items tested give evidence that the lot may include hazardous (or otherwise seriously non-conforming) products. To be both fair and effective, a mandatory sampling plan must call for enough testing to protect the consumer from hazards while also protecting both the producer and the consumer from paying the cost of unnecessary testing.

The section on cost/benefit analysis (IV) examines some illustrative major types of sampling plans in formal detail. It does this not to demonstrate the precise conditions under which this alternative is to be preferred over the others, but primarily to demonstrate a rationale for mandating in-plant sampling, supporting the conclusion that variants of this alternative should be made available to the regulator.

Alternative V

THE REGULATOR APPROVES THE DESIGN OF THE PRODUCT IN ADVANCE OF PRODUCTION VIA AN EXAMINATION AND TEST OF THE PERFORMANCE OF PROTOTYPE MODELS.

This constitutes an extensive involvement with the manufacturer in the latter's pre-production stage, and is one which strives to avoid risk at the earliest point in the production/consumption process. In many instances the manufacturer is still in the act of performing research and development on his product; moreover, this may be with respect to the determination of both its utility and its risk. Cases in point would be the approval by the Food and Drug Administration of new drug entities or by the Environmental Protection Agency of new agricultural chemicals, where research records and tests may be required to demonstrate both efficacy and toxicity. Other examples would include the approval of new automobile emission systems by the Environmental Protection Agency and of advanced design aircraft by the Federal Aviation Administration. Though most important as a means of avoiding hazard through research and test of fundamental properties of materials or systems, this alternative largely transcends the more limited scope of this report, which was to examine the role of sampling plans. Consequently, it will not be considered in further detail.

THE REGULATOR EMPLOYS A COMBINATION OF THE ALTERNATIVES.

In practice, of course, a regulator would employ more than one of the alternatives, allocating resources among them in the most effective way. Complaint-checking and support to the legal processes are necessary activities. If heavy emphasis is given to regulation of in-plant quality control and sampling inspection, the regulator would hope there would be less to do about checking products in the market place or about the investigation of injuries; however, the regulator may wish to back up mandatory in-plant testing by market place testing in order to guard against gross non-compliance due to fraud or negligence. Also the fact that the regulator may have approved pre-production prototype models does not necessarily relieve him of the need to sample and monitor production items. That relief would depend on whether the design essentially assured acceptable quality, as discussed earlier.

Alternatives I and II are particularly appropriate when the resources of the regulator are limited and must need be brought to bear on unethical and/or incompetent manufacturers who exhibit gross non-compliance. Violators are penalized with a minimum of interference with complying manufacturers.

Alternatives III and IV are most suitable for regulating safety when process control is important. In addition, of course, unsafe products are detected before they reach the market place.

Finally, Alternative V avoids hazardous design, but may need be supplemented by the other alternatives which deal with production and/or marketing stages.

Before proceeding to the discussion of in-plant sampling, it should be restated that this report does not attempt to establish precise criteria for when the regulator should use various alternatives for effecting compliance. All of the alternatives are acceptable and workable under some conditions. A number of the principal factors to be considered have been given, and are also examined formally in Section IV. It is the intent of this report to examine the viability of in-plant sampling including, on occasion, that version of it which would mandate the sampling plan along with the standard.

C. Mandatory In-Plant Sampling Plans. Given a standard and the test method for individual products, a sampling plan consists of a precription for acceptance or rejection of whole product lots by a sampling procedure. The latter may consist, for example, of requiring that, in fifteen items chosen at random from a lot*, no more than two may be found which measure below the standard level if the lot is to be accepted. In the case cited, a failure of three to "measure-up" to the standard mark of quality would be cause for rejection of that entire lot. This does not mean that the average fraction of non-conforming producs will be as high as two-in-fifteen. Should the producer set his average production quality as low as this he would find that approximately one-third of all his production would be rejected, a figure normally much too high for competitive business.

---

*The lot size is usually set by the regulator, taking into account both statistical considerations and physical constraints imposed by manufacturing and distribution practice (size of batches of raw materials, storage space, etc.).

Taken by itself the plan does not reveal the quality of products released to the public. That outcome depends on the average quality and variations around the average quality issuing from the production line itself. These in turn depend on the cost incentives of the producer as stated. The plan does have, however, a so-called "operating characteristic" or "O-C curve" which shows how the probability of accepting a lot depends on the fraction of defective* products in that lot. One point on the O-C curve of the "thirteen-in-fifteen" plan has already been given. Figure 1 displays that particular O-C curve which has been used for illustration. For example, in order to assure that at most one lot in twenty is rejected, the manufacturer must aim to produce lots having less then seven percent of defective products.

A sampling plan specifies the number of sample items to be tested, and gives the rule for judging a whole lot to be accepted or rejected on the basis of the test results. The O-C curve for a sampling plan enables a manufacturer to judge the cost of failure to maintain quality control. The regulatory agency, in selecting a sampling plan from the many available, considers all the O-C curves together with information about the technology and economics of production, and judges how much evidence of compliance is needed and how much testing should be required to forestall accidental or deliberate non-compliance.

The plan can thus be regarded as a mechanism for checking that the producer issues products of acceptable quality. It provides an incentive for his setting the quality at a level which avoids costly rejection of lots. A suitable sampling plan is selected by the regulator to achieve these objectives. As manufacturing technology and practices change, the regulator may see changing needs for evidence of compliance with standards and may alter his selection of plans accordingly.

Chapter IV on cost/benefit analysis further sorts out some of the variations in this method (Modes A, B, and C) and demonstrates how, in principle, mandated in-plant sampling in conjunction with the standard can be used by the regulator to avoid previously unaccountable damage to the health and safety of the public.

An objection raised by opponents of making a sampling plan part of the standard is that, having compiled with the protocols of the testing scheme, the producer may claim grounds for the following: a) avoiding civil liability from product injury, and b) avoiding the charge of criminal negligence by the regulator who may find non-conforming items on the shelf in his spot checking of the market place. With regard to a) the courts may often hold the producer liable for damages stemming from product use whether or not the particular product was found to be non-conforming. It is often possible, however, to make distinctions among: non-conforming products, conforming products, misuse or abuse of products by the consumer, etc., but the record of such cases in general (e.g. flammable fabrics) remains to be developed.

---

*"defective" is meant to be synonomous with "non-conforming" (see also Chapter V).

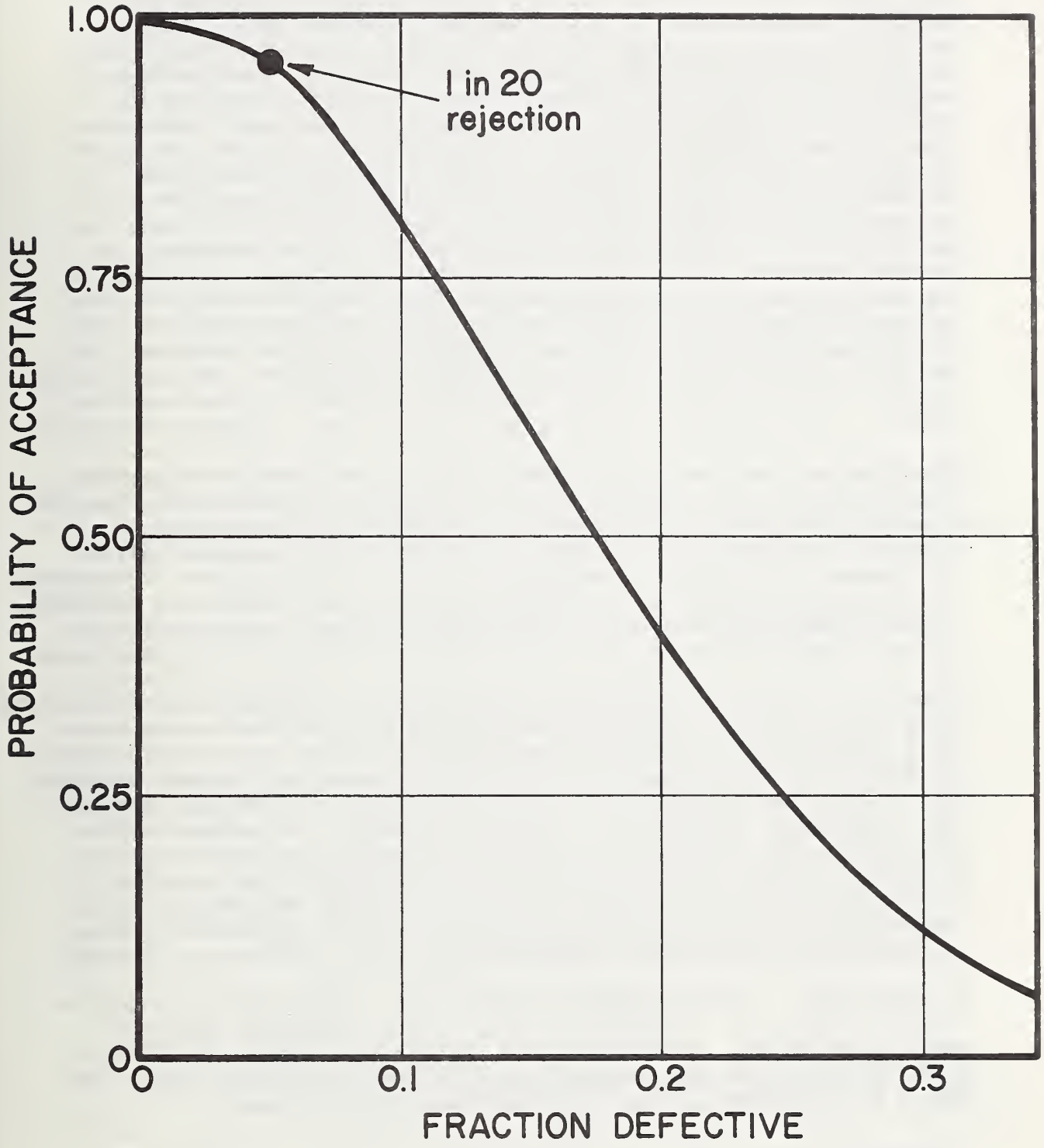


FIGURE 1: OPERATING CHARACTERISTIC FOR THE 13-IN-15 SAMPLING PLAN

With regard to b) a producer who has complied with the standard/plan should not automatically be held criminally negligent merely because one or a few non-conforming items were found in the market place. The reason for this has been given earlier; namely that no amount of costly and/or intensive effort can screen all non-conforming items from the market place. Rather than regard what may in fact be simply the reflection of unavoidable uncertainties in measurement, production, and quality control as criminal, the nature of the defect should be considered. Is the product marginally non-conforming or seriously hazardous? If the non-conforming item discovered is only marginally out of compliance, an investigation of the manufacturer's testing records and procedures may be more effective than legal action as a means to effect compliance. The legal objection to mandatory sampling plans should be met by a full realization that the plan is primarily a device which provides evidence of compliance with the standard an acceptable fraction of the time. In addition, it also constitutes an operational mechanism by which the producer observes and statistically avoids too many uneconomical rejections of product lots (see Section IV for details of this cost avoidance incentive).

Beyond the pros and cons of the desirability of utilizing sampling plans with standards there remains the issue of whether Congress in fact intended to permit their use. The opinion* of the law firm of Caldwaldder, Wickersham and Taft is that it was the intent of the Congress to allow the Consumer Product Safety Commission to employ the techniques of mandating a sampling plan as well as the standard when, in the Commission's judgement, it would constitute an effective compliance procedure. NBS cannot of course, offer expert opinions on such a matter in this or any related case.

Assuming the regulatory agency has the authority to mandate in-plant sampling plans as it sees fit, the choice would in general come down to whether or not it wished to add a sampling plan or merely "state the standard" leaving the details of the assurance testing to the producer as discussed previously under the various alternatives. Under the concept of a mandated sampling plan, compliance is sought in an orderly, automatic, and statistically satisfactory way. Avoidance of threats, incomplete ex-post examination of products requiring recall and some other inefficient and socially awkward procedures, is achieved. Mandatory in-plant sampling/inspection provides for both the manufacturer and the regulator a record of efforts to prevent the distribution of products that do not conform to a standard.

IN CONCLUSION: IN-PLANT SAMPLING PLANS CAN BE A VERY EFFECTIVE MEANS FOR ASSURING COMPLIANCE. AUTHORITY FOR THEIR USE BY REGULATORS SHOULD BE ESTABLISHED FOR HIS CONSIDERATION ON A CASE-BY-CASE BASIS. ADDITIONALLY, THERE ARE ADVANTAGES AND DISADVANTAGES TO ALL OF THE OPTIONS CITED. ANY ONE OR A COMBINATION MAY BE APPROPRIATE IN A GIVEN INSTANCE.

---

*Entered into CPSC hearings on April 4-5, 1974, by William H. Rockwell, Director of Certification, ANSI.

IV. Cost/Benefit Overview

The setting of a safety standard including the means for effecting its compliance can be regarded as a problem in cost/benefit optimization. In its simplest form cost/benefit methodology first ascertains dollar values for all of the benefits and costs involved, using appropriate probabilities to compute expected costs. A so-called "net public benefit" expression is then set up to account for the gain or benefit to society net of all costs. Included in this sum are terms which depend on the level of safety. In general, as safety is improved the cost of production increases, but at the same time the cost of accidents decreases. The optimum economic condition is usually thought to result when the last dollar invested in safety results in exactly one dollar saved in avoided accidents; however, this so called "marginal or trade-off" condition does not maximize the economic efficiency when more than one party has some control over the outcome, e.g., producer as well as regulator.

These and related matters will be presented in this chapter in an attempt to show in a general way how the standard and sampling plan operate to provide an economic incentive for the producer to comply with acceptable quality production. In addition, the exposition will show how in principle the regulator can set an optimal standard and sampling plan.

Before proceeding further, it is important to clarify the role of a hypothetical cost/benefit minded regulator in the exchange between the producer and consumer. In a free and competitive market wherein both producer and consumer are fully informed regarding the utility and safety of the product, it is a demonstrable proposition of economic theory that no need exists for interference by another party such as a government regulator. This follows from the fact that under the ideal conditions postulated an equilibrium market results wherein the producer accomodates and the consumer pays a proper price for both the utility he expects and the risks which he is willing to assume. When neither the producer nor consumer properly anticipates the hazards of a product, the producer may be held only partly accountable for the damages which ensue.* In this situation a portion of the damage cost remains outside the system of production/consumption, thus constituting what the economist terms an "externality." This is undesirable because injuries are endured which have not been anticipated, giving rise to higher costs to society as a

---

*Further complexities can arise when the product can cause damage to other members of society than to the purchaser or his immediate family, e.g., as in the driving of automobiles. In such instances, the "sovereign" right of one consumer to assume risks may interfere with his own or another consumer's "sovereign" right to avoid them.

whole than it was otherwise willing to incur. Since the producer is society's most likely agent for avoiding damage, i.e., of reducing them economically via redesign of the product, it becomes the job of the regulator, acting as surrogate for the public, to set an optimum safety standard with which the producer must comply. From another point of view the regulator's job may also be regarded as the one of effecting the missing coupling between the external costs to consumers and the internal costs to producers.

The job of this regulator or public surrogate is not to be underestimated. It requires more than the authority to set standards. It presumes also knowledge of the previously unaccountable damage to consumers, e.g., as may be revealed by epidemiology and the etiology of accidents, and additionally, knowledge of the costs which production would incur should new safety features be mandated.

In order to bring out these features in a simple but reasonable way, a type of "short run" competitive model of the firm and its industry will be invoked. Let the industry in question consist of N firms, each having a different cost function. N is assumed to be large and the production capacity of any one firm to be limited in the sense that if it attempted to produce too great an output it would eventually experience marginal costs in excess of marginal revenue.

For identical products uniform pricing must result, and for each firm this industry wide price "appears" fixed, i.e., is essentially unperturbable by the production of any one firm. If further, it is assumed that the behavior of each firm is to maximize its net revenue, then each would adjust its output to achieve that condition. Thus for the i 'th firm, the optimization of:

$$P_M q_i - C_i(q_i) \quad \text{yields (fixed } P_M) \quad \frac{\partial C_i}{\partial q_i} = P_M$$

All producers are in effect forced to sell at their marginal cost, and for the same price, P_M. The latter is not in fact fixed, but settles to a value which depends on industry's production capacity as a whole to meet the consumer demand. That is

$$\sum_{i=1}^{N} q_i = q(P_M)$$

where $q(P_M)$ is the total demand of society for the product as a function of market price, P_M. This (N + 1)st equation coupled with the N marginal cost equations permits evaluation of all of the q_i as well as the common industry price, P_M.

It is important to note that these N + 1 conditions may also be derived from optimizing the following expression, η

$$\eta = \int_0^{q_M} P \, dq - \sum_i C_i(q_i)$$

subject to the constraint: $q_M = q(P_M)$. It will be recognized that η is a type of consumer net benefit in that the integral term* constitutes a sum over "willingness-to-pay" values, P, for the product in question weighted by the consumer demand, dq. Accordingly, one may regard that the behavior of the many competing firms making up the industry has the effect of the industry group optimizing a net public benefit expression in which only those costs are included that are faced by the industry itself. It is important to note too that this is not the same behavior or effect that would result if the industry could optimize its total "collective" net revenue:

$$P_M q_M - \sum_i C_i(q_i)$$

This latter expression is rather to be evaluated from the solution to the optimization of net public benefit as defined above. The optimization of industry net revenue would require a collective effort or monopolistic collusion. In this paper the competitive condition as given above will be utilized throughout. Fixing this aspect of the problem also permits attention to be focused on safety. Also, for the remainder of this paper, the multiplicity of firms will be reduced to one "typical" firm whose production $q \simeq \frac{1}{N} \sum_i q_i$.

Thus $q_M = q(P_M)$ will in the remainder of this paper be only 1/N as large as society's total demand.

Returning now to the case where the product exhibits a safety related feature, ω, two net benefit expressions are of interest. Optimization of the first expresses producer behavior, as was shown; optimization of the second**expresses society's aim and will be taken to express the behavior of the hypothetical regulator.

---

*This term can also be written: $\int_{P_M}^{\infty} P dq$

**Termed by economists as "Pareto optimal."

$$\eta_1 = \int_0^{q_M} P dq - C_p(q_M, \omega) - C_\ell (q_M, \omega)$$

$$\eta_2 = \int_0^{q_M} P dq - C_p(q_M, \omega) - C_d (q_M, \omega)$$

In these expressions, P is the value or price which consumers are "willing-to-pay" for products at the rate of dq per unit time, whereas P_M is the actual market price. C_p is the production cost for q_M items of product (in unit time) of quality ω (with respect to safety). C_ℓ is the damage cost (over the same unit time period) for which the producer is held accountable or liable, and C_d is the actual damage cost. $C_d - C_\ell$ thus constitutes the externality. The representation is evidently for "average" or "typical" producers.

In what follows, q_M will be regarded as determined by price alone. More generally: $q_M = q(P_M, \omega)$, but for purposes of exposition we shall simplify the treatment here to the class of situations in which the consumer assumes the product is "safe enough," and thus accepts the quality of ω as given. He then demands only according to utility and price, i.e., $q_M = q(P_M)$, a function of P_M only.

When the additional feature of quality regarding safety is introduced, one sees that the unregulated producer optimizes his condition, η_1, according to:

$$\frac{\partial C_p}{\partial q_M} + \frac{\partial C_\ell}{\partial q_M} = P_M \qquad \text{and}$$

$$\frac{\partial C_p}{\partial \omega} + \frac{\partial C_\ell}{\partial \omega} = 0$$

which along with the demand schedule $q_M = q(P_M)$ determines his optimum values for P_M, q_M, and ω. The first of these conditions is the one first discussed which favors consumers under competition. The second condition expresses the trade-off cited earlier between the cost to improve quality and the cost of avoided accidents for which the producer is held liable or accountable.* Note that if the producer had been

---

*Had a monopoly been assumed the price condition would have read: $\frac{\partial C_p}{\partial q_M} + \frac{\partial C_\ell}{\partial q_M} = \frac{1 + e}{e} P_M$ where e is the so-called "elasticity of demand."

required to face C_d rather than C_ℓ, he would have automatically optimized the second net public benefit, η_2. It is essentially the regulator's job to see that this is approximated, but to do so he must have some means for bringing it about. Before this can be discussed, however, one must specify the means by which the producer himself achieves the quality level which he seeks to establish for his product.

In pursuit of the last problem posed one must first recognize that quality control is, of course, not directly affected by any end product testing, but rather through correlations which the producer has established between those tests and the various operational controls available to him in the production process. This may constitute a complex quality control scheme and could include, for example, the adjustment of pH, temperature, mechanical tolerance, and other such variables which could affect quality. Once these correlations have been learned, one might suppose one could dispense with testing. As was stated in Section III, this may be possible, but it is less likely if variations in production persist, as they will in particular when the quality sought is close to that provided by the state-of-the-art. Beyond this, records of the end product quality will be required should mandatory standards and sampling plans be instituted. Accordingly, this particular method of effecting compliance will be examined in some detail.

In general, compliance with a standard via testing or sampling will result in the rejection of those product lots which fail to "pass the test." The meaning of this last phrase will be dealt with shortly. Consider for the present that there exists some fraction, f, or overall probability of acceptance, and that this is the result of certain tests for compliance with the standard. The net benefit can then be written in such a way as to make explicit the effect of the testing:

$$\eta_1 = \int_0^{q_M} P\,dq - C_p(Q, \omega) - C_t(Q) - C_\ell(q_M, \omega)$$

subject to $q_M = q(P_M)$, as well as $q_M = f \cdot Q$. The difference between the number produced, Q, and the number released for sale, q_M, constitutes the rejects from the testing. $C_t(Q)$ represents the cost to the producer of the testing scheme itself. In general this is not negligible.

Several new quantities must now be defined. The symbol ω has been used to denote the quality level of production. In what follows ω is taken to be a numerical safety-related variable (e.g., char length) whose value can be measured for any particular item of product. The producer's action in determining quality level is equivalent to setting a mean or other characteristic value, ω_p, for the distribution of ω.

The regulator's action on the other hand may be to specify a standard value ω_s (e.g., see Mode A, below) that should be exceeded by each product item. From the distribution as may be seen in Figure 2, there is then a probability, ϕ that the measured value ω for any item exceeds ω_s. This probability ϕ depends on both the location of the distribution of quality as characterized by ω_p and on the choice of the standard value ω_s; and also on the spread of the distribution of quality which depends on the variances of both production and measurement. In other words ϕ is the probability a tested item will be found acceptable, and $(1 - \phi)$ is the fraction non-conforming or "fraction defective." The following figure illustrates these quantities.

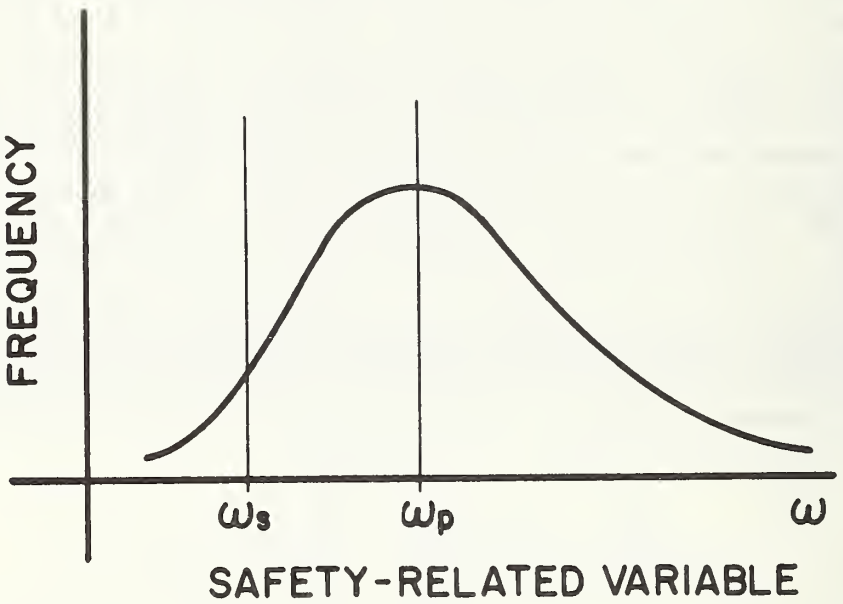


FIGURE 2: FREQUENCY DISTRIBUTION OF QUALITY, ω

The final acceptance fraction, f, is determined not only by ϕ, but also by whatever sampling rules are put into effect. In addition, it is determined by whether or not the tests are destructive of the product. To consider a simple case one may suppose that in lots of size ℓ, m samples are selected for tests and that k of these must "pass the standard" in order for the lot to be accepted. Then:

$$f = (1 - \frac{m}{\ell}) \sum_{j=k}^{m} \frac{m!}{j!(m-j)!} \phi^j (1 - \phi)^{m-j}$$

where it is assumed the tests destroy the m samples out of the ℓ products. More generally, however:

$$f = f(\omega_p, \sigma_p, \omega_s, \sigma_s, m, \ell, k, \ldots)$$

where σ_p^2 and σ_s^2 are the previously referred to variances in production and measurement, respectively, and where ... allows additional quantities to be introduced which could define more complex plans. Except for the factor $(1 - m/\ell)$ Figure 1, Chapter III, constitutes a plot of the above recipe for f as a function of $1 - \phi$ when k takes on the particular value of 13 and m takes on the particular value of 15.

For purposes of being able to characterize a variety of sampling plans and their effect on the quality of the product, it is useful to further generalize the net benefit expression to include accountability for the variances in the quantities just reviewed. Accordingly:

$$\eta_1 = \int_0^{q_M} P dq - C_p(Q, \omega_p, \sigma_p) - C_t(Q, \sigma_s) - C_\ell(q_M, \omega_p, \sigma_p)$$

subject to: $q_M = q(P_M)$ and $q_M = f \cdot Q$

This expression recognizes that costs may be associated with both the mean and the variance in the quality ω. Note that the damage term, C_ℓ is also written in this fashion, in effect stating that in addition to damage around the mean, there may be important or excessive damage at the more extreme values of the quality distribution. Not all ways of stating the standard and of the mode of testing account for this double distinction or two-parameter characterization. An attempt will be made subsequently to proceed in an orderly fashion to characterize some sampling plan types, in an elementary but sufficient way, for the purpose of revealing their implication on the mode of operations of the producer in complying with the standard. The general way of doing this is to give the solution to the optimization of η_1 and next to utilize that revealed behavior of the producer in the optimization of η_2 by the regulator. Mathematically, one carries over the set of conditions which are optimal of η_1 to the optimization problem of η_2 where they are embodied as constraints. Rather than proceed in the most general fashion invoking all of the parameters cited above, simpler net benefit expressions will first be used in conjunction with the simpler plans. This will be clarified as the text is developed.

Mode A. A STANDARD LEVEL OF QUALITY ω_s AND THE METHOD OF MEASURING ω IS SPECIFIED. THE PRODUCER IS ONLY REQUIRED TO REJECT LOTS ON THE BASIS OF SOME PRESCRIPTION, e.g., "k IN m MUST PASS."

The first thing to note about such a plan is that the acceptance fraction f is not fixed, thus permitting a variable quality product depending on how the producer optimizes his net benefit. Also this is a one parameter plan in the sense that only one benchmark, ω_s is specified in the standard.

$$\eta_1 = \int_0^{q_M} P\,dq - C_p(Q, \omega_p) - C_t(Q) - C_\ell(q_M, \omega_p)$$

$$\text{subject to:} \quad q_M = q(P_M) \quad \text{and} \quad q_M = f(\omega_p, \Sigma)\cdot Q$$

where Σ stands for ω_s and the prescription given by the sampling plan.

So far as the producer is concerned, he regards the standard and the plan as a set of boundary conditions and proceeds to optimize η_1, again only recognizing the real costs* for which he is accountable.

This gives rise to:

$$\frac{\partial C_p}{\partial Q} + \frac{\partial C_t}{\partial Q} + f\frac{\partial C_\ell}{\partial q_M} = f\cdot P_M \qquad \text{and}$$

$$\frac{\partial C_p}{\partial \omega_p} + \frac{\partial C_\ell}{\partial \omega_p} = Q\frac{\partial f}{\partial \omega_p}\left(P_M - \frac{\partial C_\ell}{\partial q_M}\right)$$

which together with the constraints yields P_M, q_M, Q, ω_p optimal for η_1.

Again the first of these conditions is the classic economic one of assuring fair price, and the second expresses the trade-off between the cost to engineer quality and the cost of avoided accidents for which the producer

---

*Note that the presence of a regulatory agency may change the function, C_ℓ so that it is different from the cost-of-damage function considered by an unregulated producer who optimizes his net benefit.

is liable; and finally, in addition, the cost of products whose rejection has been avoided by the incremental improvement in acceptable lots.* The producer, given the freedom of the Mode A prescription, will set his production quality ω_p as close to ω_s as he can, but not so close as to suffer excessive rejects. The term:

$$Q \frac{\partial f}{\partial \omega_p} (P_M - \frac{\partial C_\ell}{\partial q})$$

expresses this latter cost to the producer.

A particularly simple interpretation of this term may be made for the case when $C_\ell = 0$. Trade-off as ω_p is increased under these conditions becomes:

$$\Delta C_p = Q P_M \Delta f \equiv P_M \cdot \Delta Q$$

which in effect states that the marginal cost ΔC_p to produce improved quality should just equal the marginal value of avoided rejects (ΔQX price).

To continue with the action expected of the regulator, one must now carry over the four conditions on the optimization of η_1 to the new optimization problem of the regulator, namely of η_2:

$$\eta_2 = \int_0^{q_M} P dq - C_p(Q, \omega_p) - C_T(Q) - C_d(q_M, \omega_p)$$

subject to the four constraints cited. Note that the testing cost is

---

*It must be emphasized that in the above description of the producer's response to the standard it has been assumed that rejected lots constitute a total loss. If this should not be true, either because the producer could sell rejects to another jurisdiction, or because he could salvage an important fraction of the rejected product value, then those cost savings should have been included in the representation of η_1. Since we have presumed knowledge on the part of the hypothetical regulator about all of the producer's costs he need only recognize these to include them; however, in some practical situations these savings might not be known, and the effect of Mode A would then fall short of its design objective.

now written "C_T" to signify the inclusion of inspection costs borne by the regulator. In optimizing η_2 the regulator may vary Σ i.e., the standard conditions (ω_s, and the prescription of the sampling plan) to find out what values of the standard optimizes η_2, while taking into account the full behavior of the producer in responding to a standard and plan. It is by this two-stage process that the regulator can set the standard and plan required to couple the producer to the otherwise unaccountable damage costs.

To complete the logic of this presentation, the results of that second optimization process, namely of η_2 for Mode A should be given; however, they are avoided in this paper because of their length and complexity. They constitute a set of five equations which when combined with the four constraints yield the five variables: P_M, q_M, Q, ω_p and Σ optimal of the net public benefit of η_2. The representation Σ is evidently highly formalized, since it involves the sampling plan prescription which is not a numerical variable.

Aside from highly specific details, the final set of optimal conditions exhibit some general properties in cost/benefit optimization which arise when the regulator has limited control over decisions affecting the public interest. For example, he is unable to achieve:

$$\frac{\partial C_p}{\partial \omega_p} + \frac{\partial C_d}{\partial \omega_p} = 0$$

but only a marginal condition that must exceed zero, both because sampling is employed, and because the producer cannot be made to face the full damage cost. This lack of identity between the producer and regulator gives rise to a "best" but still imperfect accommodation of the externality.

Mode B. A STANDARD LEVEL OF QUALITY ω_s AND THE METHOD OF MEASURING ω IS SPECIFIED. THE PRODUCER IS REQUIRED TO REJECT LOTS ON THE BASIS OF SOME PRESCRIPTION, e.g., "k IN m MUST PASS." IN ADDITION, A MINIMUM FRACTION f_o OF THE LOTS TESTED MUST BE FOUND ACCEPTABLE.

This mode appears more restrictive than Mode A in that the minimum value of ω_p is ω_o, given by:

$$f_o = f(\omega_o, \Sigma)$$

where Σ is again fixed by the regulator. Whether or not the producer establishes ω_p at a still higher value depends on whether ω_o is less or greater than the value of ω_p obtained by optimizing η_1 when f can be regarded as a variable (Mode A). If ω_o is less than this then the regulator seeks the optimum standard as in Mode A. If ω_o is greater than this the regulator only embodies the constraint:

$$\frac{\partial C_p}{\partial Q} + \frac{\partial C_t}{\partial Q} + f_o \frac{\partial C_\ell}{\partial q_M} = f_o P_M$$

along with $q_M = q(P_M)$ and $q_M = f_o Q$ in the optimization of η_2. The details will not be given here, but it can be shown that the ultimate optimum condition arrived at by the regulator is the same in both cases, i.e., that the solution to Mode B is for the regulator to select f_o such that ω_o has the same value as that value of ω_p which would have optimized Mode A. In actual practice however, this value may not be known and a real regulator may prefer to fix f via the Mode B prescription.

A certain difficulty exists for the regulator in effecting compliance through the Mode B approach. This has to do with the fact that only time can reveal whether the fractional acceptance of lots exceeds the minimal value of f_o or not. There thus exists an ambiguity regarding the action the regulator should take following various kinds of accumulated evidence of "probable" non-compliance. A related ambiguity faces the producer in seeking to optimize his net benefit over time. One way of dealing with this situation is to operate under two plans: a strict and a relaxed version. Additional "switching rules" are set up to determine whether the producer is allowed to operate under the relaxed plan or under the strict plan. In general, if the producer's history under the strict plan gives "proper" evidence of compliance, he is allowed to shift to the relaxed plan and vice versa. The switching rules determine when these options arise; however, these modes of operation are too complex to be presented here. See also Chapter V. They have the virtue of hedging against the uncertainty of exactly how the producer will respond. Some of the acceptance sampling procedures used in military procurement are similar in spirit to Mode B (see Chapter V).

In the discussion above of Modes A and B, changes in production quality were represented as changes in a mean or characteristic value ω_p selected by a producer, and the relation between ω_p and the fraction non-conforming $(1 - \phi)$ was indicated in Figure 2. Thus, it was implicitly assumed that the shape of the distribution of ω, in particular its variance, was the same for any mean value. If this were not so and the producer could avoid rejection of lots by narrowing the distribution

through tighter quality control while at the same time keeping the mean quality itself fixed, his response to Modes A and B would be different. To allow for this in part, most mathematical analyses of the performance of sampling plans that make a "pass-fail" test are done in terms of the fraction non-conforming. To allow for changes in both mean and variance in the cost/benefit analysis, it would be necessary to consider including both the mean and the variance in the production cost and damage functions.

In the next and final mode to be discussed in this section, consideration is given to a two-parameter standard; not, it is assumed, because the producer has practical control over the spread in quality independent of the mean value, but rather because it is assumed the damages are significantly greater when the product quality falls below some critical value designated as ω_s.

In addition, the desirable level of quality is specified by a second benchmark, $\omega_o > \omega_s$, with a separate and less stringent sampling prescription.

Mode C. TWO STANDARD LEVELS OF QUALITY ω_s AND ω_o AS WELL AS THE METHOD OF MEASURING ω ARE SPECIFIED. THE PRODUCER IS REQUIRED TO REJECT LOTS ON THE BASIS OF TWO PRESCRIPTIONS, e.g., "k_s IN m MUST PASS ω_s," AND "k_o IN m MUST PASS ω_o."

In this scheme measurements are made as before on m samples, and note is made as to whether a particular item passes ω_s, but not ω_o or whether it also passes ω_o. Products thus fall into three classes: $\omega < \omega_s$; $\omega_s < \omega < \omega_o$; and $\omega_o < \omega$. Different damage costs are associated with these classes, and probabilities need be utilized which determine the number of products present in each of these classes.

No attempt will be made to further analyze this complex case in detail; however, the producer again sets ω_p to compromise between C_p and these several level damage costs. In order to determine the optimum regulator response, one would again carry the producer's response over to the regulator's problem, where they would appear as constraints in the optimization of η_2. Two benchmarks and two prescriptions for sampling are set by the regulator wherein: $k_o << k_s$. (Chapter V gives a reference to recent work on such "three-class attributes" sampling plans.)

In modes analogous to B the regulator could also establish minimal probabilities for passing either ω_s or ω_o. These will not be analyzed either, though similar but somewhat more complex results than those obtained before would ensue. Also as before, switching rules and alternate plans may be advantageous to employ.

Evidently there are more complex modes which could be pursued in the fashion outlined, particularly ones which recognize independent control over the variance in the quality of production, σ_p^2; however, the above should serve to illustrate the general method which in principle could be followed in the setting of standards and of related sampling plans or systems to effect compliance. In practice, of course, much of the implied information is either not known or only poorly known. Nevertheless, it is important for the real regulator to be aware of what is involved, so that when best "guesstimates" and judgment of "reasonability" are made, the fundamental factors are not neglected. The exposition given purports to have identified what those factors are.

A fuller treatment of cost/benefit analysis under uncertainty, (i.e., incomplete knowledge) wherein the costs to perform research and development to reduce uncertainty would also be included in a dynamic version of the above is beyond the scope of this report. In general, if uncertainties are too large no immediate decision may be called for, except perhaps the performance of research and development. This might be directed for example, to the improvement of knowledge of either epidemiology, or of the production cost functions. The decision by the regulator to effect a standard might in some instances be based only on the estimate that $C_d >> C_\ell$ and that net beneficial improvements in the quality ω_p could be effected without great increases in cost, C_p. If so, the regulator then seeks to determine what "reasonable" improvements in safety could be made, i.e., what ready technology might be invoked to avoid the worst accidents. Having made such a determination, he would then wish to set the standard and the sampling plan or system to bring about the desired result. He might choose to do this in two stages, wherein the first stage would be used to reveal improved details for application to the second stage of standard and plan (system). Under uncertainty these could hardly constitute optional maneuvers as was posited earlier to be the aim of the hypothetical regulator. Under uncertainty a real regulator might be content if his setting of a standard simply constituted some probable net beneficial gain to society.

Another "testing" method for effecting compliance is one in which the regulator "spot checks" products found in the market place itself; i.e., from "off the shelf" combining Alternative II with in-plant sampling inspection.

In effect, market sampling plans which ride "piggy-back" on in-plant plans constitute further incentive to the producer to strive for quality. The cost/benefit method of analysis could again prescribe the optimal hybrid arrangement between the two sampling regimes, but is much too complex to be analyzed here; moreover, little development of plans of this type has been carried out to date.

Though in some instances market sampling and testing in adequate depth may be feasible, it can in many other cases become an enormous burden and therefore be cost-ineffective for the regulatory staff to undertake. There exists a point of view in this regard which maintains that the regulator need not sample often nor for all producers because the "threat" that he may is enough to effect compliance. This argument in the long run is at variance with a cost/benefit point of view which seeks to establish socially desirable goals through identifiable rational incentives. Some brief comments on the cost/benefit considerations follow.

If the regulator would attempt to "make-up" for infrequent testing by imposing large penalties whenever defective products are found, the system of production might find itself subject to excessive fluctuations, again a socially undesirable feature. It is preferable to "make small punishments fit small crimes." Though the exact rationale for this will not be undertaken in this paper, suffice it to say that evidence should be accumulated in a time scale short compared with the overall life-time of the production/consumption process. In this way large gambles are avoided and the public is assured of a steady acceptable quality. A deeper involvement in the quality control measures of the producer is also possible and in the case of military procurement such practice is often followed. No attempt will be made in this paper to state precise criteria for when such practice may be feasible. For further comment see Chapter III of this report.

One may, of course, always entertain the notion that the record may be fraudulent thus requiring some further means of testing by the regulator. Spot checks followed by in-depth sampling in the market place can be used for this or any other suspect case. Once the producer has adhered to a prescribed testing procedure, however, one may assume that the sought for coupling to assure compliance has been realized. It should be appreciated though that this assumption is ad hoc to cost/benefit analysis. Its validity rests on other experience, namely the general one that if a producer's records are open for inspection, whether this be for disclosure of finances or for gaining evidence as to the quality of production, the producer will likely find the strategy of fraud unrewarding. Under such conditions, too, it may be legitimate to treat fraud as an event whose penalty may exceed the crime, thus further inducing the producer to follow the rules laid down by the administrative law of the regulator.

IN CONCLUSION IT MAY BE STATED THAT FROM THE COST/BENEFIT OVERVIEW OF THE SUBJECT THE USE OF SAMPLING PLANS IN CONJUNCTION WITH STANDARDS MAY CERTAINLY BE AN EFFECTIVE MEANS FOR ASSURING COMPLIANCE, AT LEAST BY HYPOTHETICAL REGULATORS WHOSE AIM IT IS TO SET STANDARDS OPTIMAL OF THE NET PUBLIC BENEFIT AS DEFINED BY COST/BENEFIT METHODOLOGY. ACCORDINGLY, IT WOULD APPEAR DESIRABLE TO MAKE THIS TECHNIQUE AVAILABLE TO REAL REGULATORS AS WELL.

V. The Use of Sampling Plans in Compliance Testing

Although a rationale, supported by simplified examples (Modes A, B, and C), was presented in Chapter IV whereby an all-knowing regulator could set standards and sampling plans optimal of the public interest, in practice the overall problem is divided into several major parts and separately analyzed by different disciplines. For example, it was pointed out in Chapter III that to devise a method for the measurement of performance a technical body such as NBS might be called upon by the regulator to determine an appropriate measurement technique. If completed successfully, the standard method of measurement so devised would correlate perfectly with the hazard in question. In the parlance of the previous chapter, a particular performance measurement, ω, would allow the regulator to infer the residual damage, C_d, posed by that item of product. Though such is the intent, in actual practice the measurement only imperfectly correlates with true hazard. In spite of this the regulator proceeds as if this were a precise method and next considers other imperfect information such as the cost of production and compliance. In managing these imperfectly fitting and partially decoupled regimes, the regulator may iterate a few times the output of one regime as input to another regime.

One of the major breaks in consideration of the total problem comes in the selection of an appropriate sampling plan or scheme. In this case the regulator turns to the statistician or quality control specialist to devise a suitable sampling plan or scheme for determining conformance to and providing an incentive for compliance with the standard. Though again the plan and the standard should ideally be jointly determined as envisioned in Chapter IV, the standard is arrived at "first" by a judgment of the many factors already cited, and the plan is thought of "second" as a device to check conformance with that standard. Whether this dichotomy of function is regarded as absolute or only as strategic, it leads the regulator and the statistician to confront jointly certain questions which must be answered in order for the specialist to design an appropriate sampling plan or scheme. For example, he must be informed of how important it is for various fractions of non-conforming items to escape the statistical screen and reach the market place. More generally, the selection of a suitable sampling procedure involves policy decisions about the amount and kind of evidence needed, from which statisticians can calculate the details of a specific set of sampling plans. This chapter discusses some of the technical-policy alternatives that must be considered.

A. Conformance to a Standard. As stated above, the statistician regards the standard as he does a specification, namely,as a requirement which the product should meet. The purpose of the sample checking would then be to determine whether or not those requirements have been met. In considering the role of sampling in determining conformance with a

standard, one can draw a rough parallel with the legal process. The requirements of a standard are like the law. The sampling procedure provides rules for obtaining evidence and also provides an automated judge and jury to make a decision as to whether the product is in conformance with the standard. Curtiss [1]* has discussed this distinction with regard to military specifications. Since the two functions are distinct, two basic questions should be considered:

° How is conformance to a standard to be defined?

° How is conformance to a standard to be judged?

Although the sampling procedure is to be used to judge conformance, the design or selection of the proper sampling procedure involves much more than just the selection of a few key numbers. It will depend on how conformance is defined, how inspection is to be done, what kind and degree of protection against non-conforming product is desired.

The basic question in defining conformance is: "Does the standard define conformance for an item, for a lot, or for a process?"

Traditionally, e.g., in military purchases, specifications were thought of as defining conformance for an item, and the supplier was expected to supply all items conforming. In order to obtain the many benefits of sampling inspection, a contract might specify some acceptable quality level (fraction defective greater than zero) but MIL-STD-105D [2] states that, "The designation of an AQL shall not imply that the supplier has the right to supply knowingly any defective unit of product."

Presumably most product safety standards are intended to define conformance for an item. Here is where some misunderstanding of the role of sampling can arise. When it is intended that every item conform to the standard, the requirements for an item are, in effect, translated into requirements for a lot or a process, in order to obtain the many benefits of sampling inspection.

The only conceivable way to determine conformance for every item is to check every item, but it is well known that 100 percent testing has its drawbacks. It is sometimes impossible, sometimes overly expensive, and has been demonstrated to be less than 100 percent effective due to human error. Automatic 100 percent screening (or more than 100 percent--e.g., double screening) may be possible in some cases, but has the same limitations. For destructive tests, 100% testing is of course impossible. There is, therefore, an effective translation of the requirement for

---

*Numbers in square brackets refer to the list of references given at the end of this chapter.

every item to conform into some requirement for an allowable small percent non-conforming for the lot or process. As long as every single item in the lot is not tested, it is impossible to say that the percent defective* in lot is zero.

In addition to this operational translation into an allowable percent defective for the lot in order to realize the benefits of sampling inspection, it is possible to have standards that define conformance in terms of some characteristic of a lot or process, rather than an item. Consider the following possibilities for the definition of conformance; for example:

(a) "An item is conforming if it is/does...." The detailed description specifies one or more requirements for the item. These requirements may have the nature of "design" requirements or "performance" requirements.

(b) "A lot is conforming if..." The detailed description should specify some quality requirement for the lot collectively. Some possible requirements are:

1. for the percentage of defective items in the lot, where defective items have been defined as in (a) above.

2. for the lot average of some chosen property.

3. for a measure of the lot variability for some chosen property. There are a number of other possibilities. Some of them can be used in combination, and some of them in combination imply others.

For bulk product, for example, the definition of a conforming item as in (a) would have been impossible, and traditionally a specification or standard has specified some property of the lot or batch--most often the average.

Note: Some people erroneously use a sort of operational definition of a conforming lot as a lot which passes the specified sampling procedure. (When conformance is instead defined in the ways given above, and if sampling is used, a non-conforming lot may pass and a conforming lot may not pass.) This erroneous definition may be attractive, but it begs the question of how conformance is or should be defined and how conformance should be judged.

---

*A "defective" is simply an item not conforming with the standard. Because the word sounds bad, some efforts are being made to use another word. However, the words "defective" and "percent defective" are classic in the literature of acceptance sampling, and will be used here.

(c) "A process is conforming if...." A conforming process could be defined in terms similar to "design requirements" for an item, i.e., certain specified physical characteristics of the process,* but this is not discussed here. Instead we consider only quality requirements, i.e., certain collective properties of the process, parallel to those mentioned for a conforming lot in (b) above. Ordinarily, one would require that the process be in statistical control with a specified limit(s) on:

1. the process average in percent defective (items)

2. the process mean of some chosen property

3. the process variability for some chosen property.

In the matter of judging conformance, the first consideration is the kind of definition of conformance. Certain procedures involving sampling are appropriate to one or the other of the definitions discussed above. There are a number of other considerations that influence or dictate the choice of sampling procedure. The list might include the following:

(a) Does the standard define conformance for an item, for a lot, or for a process?

(b) Is the objective to accept/reject individual lots of product or is the intention to require/monitor the control of some process?

(c) Is inspection or testing to be applied to a series of defined aggregations of product (e.g., lots) or to a continuous stream of product?

(d) What is the nature of the unit which is to be inspected or tested? One kind of unit (and probably the most common in the product safety area) is a discrete item of product. Other possible units are arbitrarily defined units (such as a yard of fabric from a bolt, a composite sample from a bulk product, a well-defined test specimen for a particular test, etc.).

(e) What kind of observation is made as a result of the inspection or test on a unit? One possibility is a simple classification into one of two categories. (Some very recent work has been done with regard to a three-category classification scheme). Another possibility is the actual measurement of some property.

---

*This may be done in regulation of food processing, for example.

When a measurement is taken, however, the actual value of the measurement may be used or it may be used only to obtain a two-category classification such as above or below some specified value.

(f) What kind of statement is desired or can be made about the sampling procedure as a means of checking or enforcing the requirements of the standard? What kind of assurance is afforded by the sampling procedure?

Question (a) specifically relates to how one defines conformance. We noted that we must first decide whether the standard defines conformance for an item, a lot, or a process. If the standard defines conformance for a lot or a process, it is concerned with a collective property of the lot or process. We might now consider these in some more detail.

Some examples of lot or process quality requirements:

1. for the lot or process percent defective

2. for the lot or process average of some property

3. for the lot or process variability.

These are quality characteristics that might be important in the use of the product in the real world (the percentage of non-conforming items, the lot average for bulk product, etc.). Which one of these lot or process quality requirements would be used to define conformance would depend on what is meaningful. It will be recognized that the quality requirement on percent defective is the most popular one, because it involves no assumptions about the distribution of a measured property. It therefore has been the real work horse of acceptance sampling. Nevertheless, sometimes the mean value of some property is the requirement of concern to the users; sometimes the variability. Furthermore, it may be desirable in compliance testing to consider quality requirements that do involve the distribution of the relevant quality characteristic, for example, to limit the variability of the distribution so that items that depart from the intended quality would be at worst only marginally non-complying.

Questions (b) through (f) also influence how conformance is to be judged and are discussed below.

Question (b)--Accept/reject vs. control of a process. Acceptance/rejection of individual lots of products uses statistical tools and other procedures which come under the general heading of "acceptance sampling inspection." Specifying or monitoring the control of a

process would involve the use of statistical control charts. Some systems of acceptance sampling inspection do result in the acceptance/rejection of lots, but have features which operate to exert control on a process in some way.

Question (c) Lot-by-lot inspection vs. inspection of a continuous stream of product.

Not all types of acceptance sampling procedures operate lot-by-lot. Some are designed for a moving product and are called continuous sampling plans.

Question (d) Nature of unit inspected or tested. In the interest of simplicity, most of this discussion is framed in terms of an "item," i.e., a discrete unit of product. Bulk product poses special problems which are not discussed here. Other special cases involve the unit which is an arbitrarily defined length or area (e.g., a yard of cloth from a bolt, some length of wire from a reel). Where a test is required to be made on a well-defined test specimen, which is not the same as an item, inferences are made by imagining that the whole lot (for example) is cut up into such specimens.

Question (e) What kind of observation is made on an item or unit? A very common observation is to classify into one of two categories--e.g., pass-fail, go-no go, above or below some limit. This is commonly called inspection by attributes. (It might now be called "two-class attributes" because there are some very new sampling plans for three-class attributes inspection, e.g., a classification scheme of Good-Marginal-Bad [3]).*

Another kind of observation is to make a measurement of some property. This is called inspection by variables. Of course, it is possible to make this measurement but to use it only as in two-class attributes, i.e., to decide whether it is above or below some limit. When the measurement is actually used, it may be used to compute some statistics from the sample (e.g., sample mean and sample estimate of variability). This does not imply that the lot or process requirement is necessarily defined as lot or process average or variability. The sample information can be used as an estimate of lot average or variability or can be used in connection with a requirement for lot or process percent defective (if the necessary distribution assumptions can be made).

---

*See also discussion of Mode C, Chapter IV.

Another kind of observation is to count the number of defects per unit (e.g., in the case of visual inspection of textiles).

Question (f) What kind of statement is desired or can be made about the sampling procedure as a means of checking conformance with a standard? What kind of protection is afforded by the sampling plan? This will be discussed in detail in B below.

At this point we might indicate how certain well-known sampling schemes fit into this framework. Take for example MIL-STD-105D, (now also ANSI Z1.4) "Sampling Procedures and Tables for Inspection by Attributes" [2]. The use of MIL-STD-105D implies: (a) process requirement in terms of percentage of defective items (its use for so-called defects per unit inspection is not discussed here), (b) acceptance/rejection, (c) lot-by-lot inspection, (d) discrete units, (e) classification of unit as defective/nondefective.

MIL-STD-414, (now ANSI Z1.9). "Sampling Procedures and Tables for Inspection by Variables for Percent Defective," [4] implies the same kind of process requirement, i.e., percentage of defective items, also applied to the acceptance/rejection of lots. Instead of a 2-class classification of a unit, however, a measurement is made. These measurements and calculations from them are translated into estimates of percent defective. If in fact the requirement is for a lot or process average, MIL-STD-414 could not be used. (For this case, see e.g., ASTM E-122-72 [5], or the book by A. J. Duncan [6]).
The large available collections of sampling plans are for the percent defective type requirement for lot or process. MIL-STD-105 and 414, for example, differ only in regard to the kind of observation made on the item.

MIL-STD-105 and 414 have another feature in common--the feature raised in Question (f) having to do with the type of assurance afforded by the sampling procedure. This is a more complicated feature and is discussed in B, but first a discussion of the difference between a sampling plan and a sampling scheme is in order.

B. Sampling Plans and Schemes. A sampling plan is a statement of how many items are to be inspected or tested and exactly what decision is to be made based on the results of such inspection or test. A sampling scheme according to Hill [7] is an overall strategy specifying the way in which sampling plans are to be used. He notes that "the study of sampling plans may be thought of as a purely mathematical exercise," but that the study of sampling schemes is "a matter of art, opinion, aesthetic sense and compromise as well as of science and mathematics."

There are several well-known sampling schemes, e.g., MIL-STD-105 and 414, Dodge-Romig. Each contains a large set of sampling plans, and also contains rules for the operation of the scheme, with the aim of achieving some overall quality objective. Hill's paper discusses features of a number of such schemes.

The assurance afforded by the use of a particular sampling plan is given directly by the operating characteristic curve of the sampling plan, which shows the probability of accepting lots as a function of lot quality. A sampling scheme is described in terms of how it aims to provide assurance about quality. The major sampling schemes have various objectives in this regard, including the three which will be explained below:

AQL schemes (e.g., MIL-STD-105D [2] and MIL-STD-414 [4])

LTPD schemes (Dodge-Romig [8])

AOQL schemes (Dodge-Romig [8])

AQL schemes. AQL schemes aim to provide assurance about the process average percent defective through the economic pressure of acceptance and rejection of lots. AQL means "Acceptable Quality Level," which is defined in [9] as follows:

> "Acceptable quality level (AQL): The maximum percent defective (or the maximum number of defects per hundred units) that, for purposes of acceptance sampling, can be considered satisfactory as a process average.*
>
> *When a consumer designates some specific value of AQL for a certain characteristic or group of characteristics, he indicates to the supplier that his (the consumer's) acceptance sampling plan will accept the great majority of the lots that the supplier submits, provided that the process average level of percent defective in these lots is no greater than the designated value of AQL. Thus, the AQL is a designated value of percent defective (or of defects per hundred units) that the consumer indicates will be accepted a great majority of the time by the acceptance sampling procedure to be used. The AQL alone does not describe the protection to the consumer for individual lots but more directly relates to what might be expected from a series of lots, provided that the steps called for in the reference AQL system of procedures are taken. It is necessary to refer to the OC curve of the sampling plan that the consumer will use, or to the AOQL of the plan, to determine what protection the consumer will have."

The schemes act to apply pressure on the producer to produce lots of acceptable quality. If he does not, he suffers increasing pressure (in the form of more frequent rejections) from application of the rules

of the scheme (e.g., tightened inspection).* However, these schemes do not protect the buyer in direct fashion and they assume that every lot is inspected in a continuous stream of lots.

LTPD schemes. LTPD (Lot Tolerance Percent Defective) is defined in [9] as follows:

> "Expressed in percent defective, the poorest quality in an individual lot that should be accepted. Also referred to as Rejectable Quality Level (RQL)."

The sampling plans used will be characterized by a low frequency of acceptance for lots of LTPD quality or worse. They are now considered to be more appropriate for the situation of sampling an isolated lot.

However, the Dodge-Romig tables [8] contain an LTPD sampling scheme, designed to minimize the amount of inspection under certain assumptions.

AOQL schemes.** An AOQL (Average Outgoing Quality Limit) scheme [8] aims to control average outgoing quality by screening (i.e., removing all defectives from) rejected lots. Again the assumption is made that all lots are inspected. If it is possible to screen all rejected lots (a fairly restrictive assumption in the real world and an impossible one for a destructive test), a value of average outgoing quality can be calculated for various values of as produced quality, and the worst average outgoing quality can be determined. This is called the AOQL (Average Outgoing Quality Limit).

More discussion of all these schemes will be found in [6], [7], and [10].

In summary, the large collections of sampling plans (e.g., the attributes sampling "schemes" such as MIL-STD-105, Dodge-Romig, etc.) differ not only in their individual sampling plans and in the rules for using them, but in what they intend to achieve and how it is achieved. Each scheme offers choice in regard to the degree of assurance, but each is different in regard to the basic kind of assurance provided.

C. Use of Sampling in Compliance Testing. The different kinds of sampling schemes (LTPD, AOQL, AQL schemes) have been recognized to be suitable for different purposes and different buyer-seller relationships. AQL schemes were originally developed for use in military procurement, and have also proved to be popular with large industrial buyers. The compliance testing situation is yet a different one, and probably needs a new kind of scheme.

---

*See also discussion of Mode B, Chapter IV.

**Sampling plans used to inspect a continuous stream of product (called Continous Sampling Plans--CSP) are also characterized by AOQL.

In particular, several features of AQL schemes should be carefully considered before these schemes are applied in compliance testing. First, they assume that every lot is inspected in a continuous series of lots from the same producer. The economic pressure exerted by AQL schemes to keep the producer at the AQL or better is just not there when only occassional lots are inspected, and they are obviously not appropriate for market place checking.

Second, even when used as designed, AQL schemes achieve their aim in a fairly complicated way. For example, Gerald J. Hahn and Edward G. Schilling [10] have said:

> "MIL-STD-105D is not simply a collection of sampling plans. Rather, it is a sampling scheme. As such, its operation combines several individual sampling plans into a procedure designed to use economic, psychological, and operational means to motivate the producer to sustain quality at least at the level of a prescribed AQL. The switching rules are at the heart of this procedure. Correct application of MIL-STD-105D implies full use of the scheme it presents."

Some elementary form of economic incentive is present in any sampling procedure which rejects whole lots rather than individual items. The economic incentives of AQL schemes are more specific and sophisticated, but exactly how well they work is not really known. One of the few evaluations of these schemes is contained in a paper by K. S. Stevens and K. E. Larson [11].

Third, AQL schemes constituted a direct statement to a seller from a buyer on how the specification requirements were to be enforced. Emphasis was given to the treatment of acceptable products. A regulatory agency, acting on behalf of a large number of individual buyers, must also make some statements to them emphasizing how the buyer is protected from receiving unacceptable products. The analysis needed to support such statements would include studies of the impact of sampling schemes on non-complying producers. More investigation of the operational factors in, and possible underlying models for, the compliance testing situation is needed.

In summary, the sampling plan used for illustration in Chapter III is one of the simplest kinds of sampling plans. It is also a special kind of plan that presupposes that a number of basic decisions have been made about how sampling is to be used to determine conformance to a standard. Various kinds of sampling plans are available for use in developing procedures, called sampling schemes, for compliance testing. The existing sampling schemes were devised for large-scale procurement, primarily military. Sets of sampling plans were put together into schemes or systems characterized by some desired overall objective in the way of quality assurance. There has been little experience to date

to demonstrate the effectiveness of sampling schemes for compliance testing. Such experience, and related mathematical investigations, are needed for the formulation in general terms of the overall objectives of sampling schemes, so that the statistician and the regulator--given the standard--can make and explain an appropriate selection.

UNTIL A VARIETY OF GENERAL SCHEMES HAS BEEN DEMONSTRATED FOR UTILIZING SAMPLING PLANS IN COMPLIANCE TESTING, SAMPLING PROCEDURES MUST BE DESIGNED ON A CASE-BY-CASE BASIS.

Recommended Reading

The history of the development and use of acceptance sampling plans from their beginnings in the Western Electric Co. in the 1920's is described in a series of articles by Harold F. Dodge in the Journal of Quality Technology [12].

Sampling Inspection [13], by the World War II Statistical Research Group at Columbia University, contains very good basic material on acceptance sampling plans. Because the definition of AQL given there is no longer used, it is out of date as a sampling scheme. A discussion of sampling plans and sampling schemes and the philosophy of various sampling schemes is given by Hill [7]. One of the few evaluations of how AQL schemes work to keep the producer producing at the AQL is found in [11]. The best book on how-to-do-it for designing sampling plans (as well as much theoretical discussion) is Duncan [6]. Useful Guides to MIL-STD-105D-type sampling schemes are contained in Australian Standard 1399-1973 and in ISO 3319, Guide to the Use of ISO 2859. An ASTM Special Technical Publication in preparation, Selected Papers on Acceptance Sampling, will have a very helpful Introduction.

References

[1] John H. Curtiss, "The Application of Statistical Theory to the Preparation of Industrial Specification and Acceptance Procedures," Proceedings of the International Statistical Conferences, 1947, vol. III, pt. A, 351-357.

[2] MIL-STD-105D, Sampling Procedures and Tables for Inspection by Attributes, Department of Defense, 29 April 1963.

[3] D. F. Bray, D. A. Lyon, and I. W. Burr, "Three Class Attributes Plans in Acceptance Sampling," Techometrics, vol. 15, no. 3, August 1973, 575-585.

[4] MIL-STD-414, Sampling Procedures and Tables for Inspection by Variables for Percent Defective, Department of Defense, 11 June 1957.

[5] ASTM E-122-72, Standard Recommended Practice for Choice of Sample Size to Estimate the Average Quality of a Lot or Process.

[6] Acheson J. Duncan, Quality Control and Industrial Statistics, 4th Edition, Richard D. Irwin, Inc., Homewood, Illinois, 1974.

[7] I. D. Hill, "Sampling Inspection and Defence Specification DEF-131," Royal Statistical Society, A, vol. 125, pt. 1, 1962, pp. 31-87.

[8] Harold F. Dodge and Harry G. Romig, Sampling Inspection Tables: Single and Double Sampling, 2nd Edition, John Wiley and Sons, New York, New York., 1959.

[9] American Society for Quality Control, Glossary and Tables for Statistical Quality Control, available from the ASQC, 161 W. Wisconsin Avenue, Milwaukee, Wisconsin 53203.

[10] Gerald J. Hahn and Edward G. Schilling, Private Communication.

[11] K. S. Stephens and K. E. Larson, "An Evaluation of the MIL-STD-105D System of Sampling Plans," Industrial Quality Control, vol. 23, no. 7, January 1967, pp. 310-319.

[12] H. F. Dodge, "Notes on the Evolution of Acceptance Sampling Plans," Journal of Quality Technology:

Part I. vol. 1, no. 2, April 1969, pp. 77-88;
II. vol. 1, no. 3, July 1969, pp. 155-162;
III. vol. 1, no. 4, October 1969, pp. 225-232.

[13] H. A. Freeman et al., Sampling Inspection, McGraw-Hill Book Co., Inc., New York, New York, 1948.

APPENDIX A

SETTING A STANDARD

Ian R. Bartky

Introductory Note

This material is taken from a larger background paper prepared in April 1974 for staff members of the U.S. House of Representatives Committee on Interstate and Foreign Commerce. It was written to provide a very basic introduction to statistical sampling plans and their association with the standards development and enforcement process. The section of it given below describes how a decision-maker might set a level in a standard for a class of items in commerce based upon a survey of the state-of-the-manufacturing-art. A few modifications to the original section have been made here.

Setting a Standard

Consider a case where we decide to establish a standard based on some test method. (We're not considering standards which might be set on, say, the basis of medical criteria.) We examine what is available in the market and we perhaps find there are several manufacturers using different processes which give different results in our tests. (Assume the test methods themselves are causing no problems.) Our results are shown in Figure 1 which indicates the frequency distribution of the relevant quality property for five manufactures.

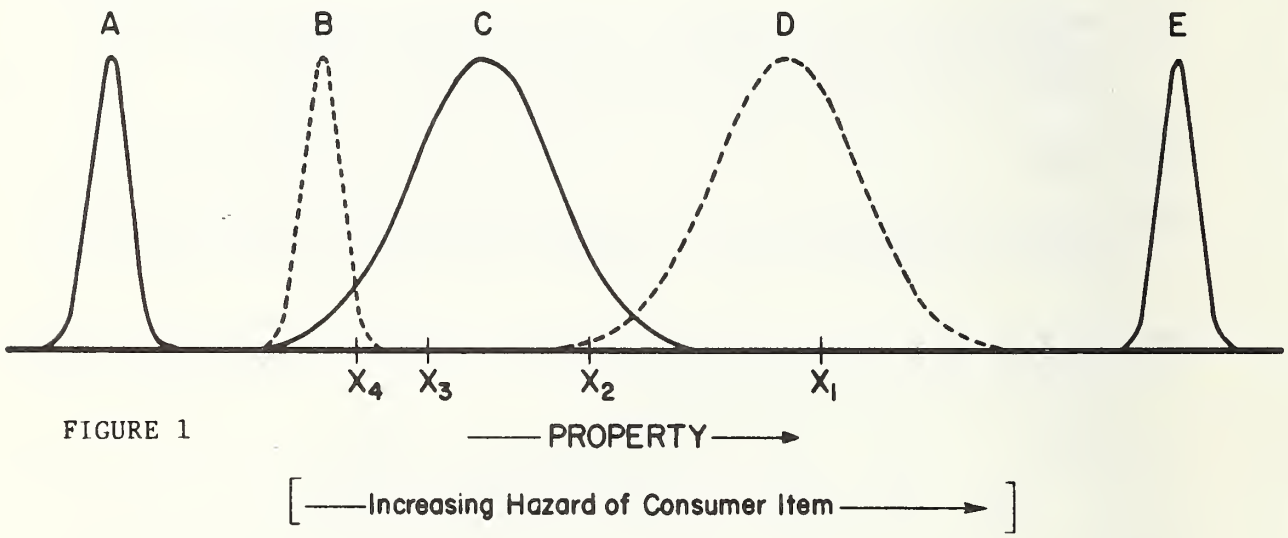


FIGURE 1

We might decide that the quality level of manufacturer E's items is unacceptable in the market (e.g., a "torch" sweater). Thus we have no qualms about setting a level which will exclude his production.

We find that although the consumer safety implied by the sole sale of items made by manufacturer A is the highest, the process (or raw materials, or cost) is so unique that no other manufacturer could change his process to achieve A's level of safety and remain in business. Consequently, we decide we cannot set the level to exclude B, C, and D.

We learn that B, C, and D are not very much different in their processes. With a little bit of effort (cost), D could increase his quality level so he mirrors C's level of safety (i.e., D's curve could be shifted so it lies upon C's): Figure 2.

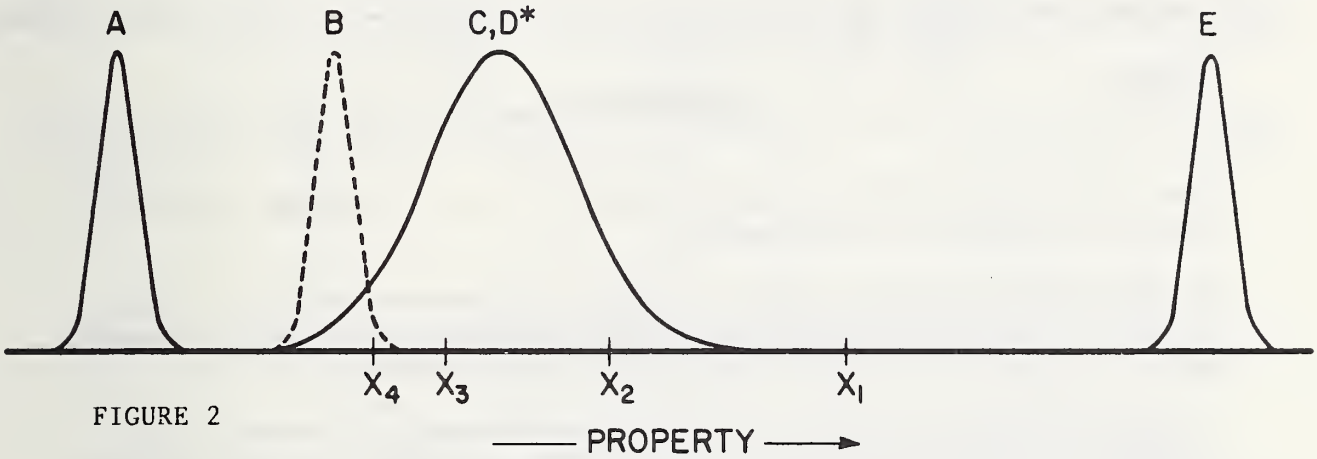


FIGURE 2

We also find that C's quality control could be upgraded significantly so that his distribution would be as good as as B's (i.e., as narrow as curve B), and this requires only modest cost increases: Figure 3.

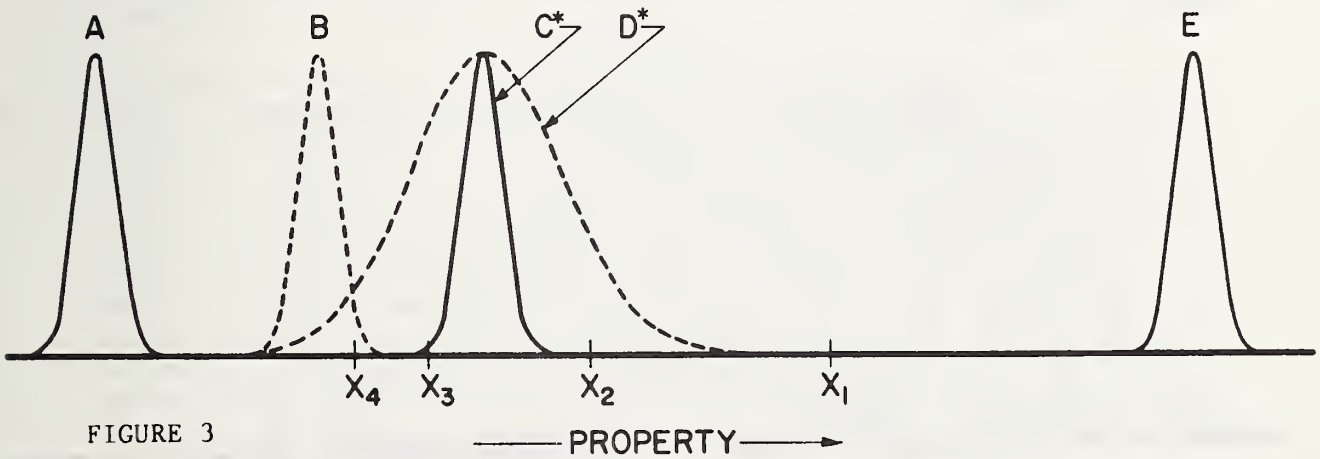


FIGURE 3

We still know that C and D could improve their items even farther so that their production would have the same level of quality as B (i.e., shift both their curves to lie on B) and would have the same quality range as B (i.e., all three have narrow curves): Figure 4.

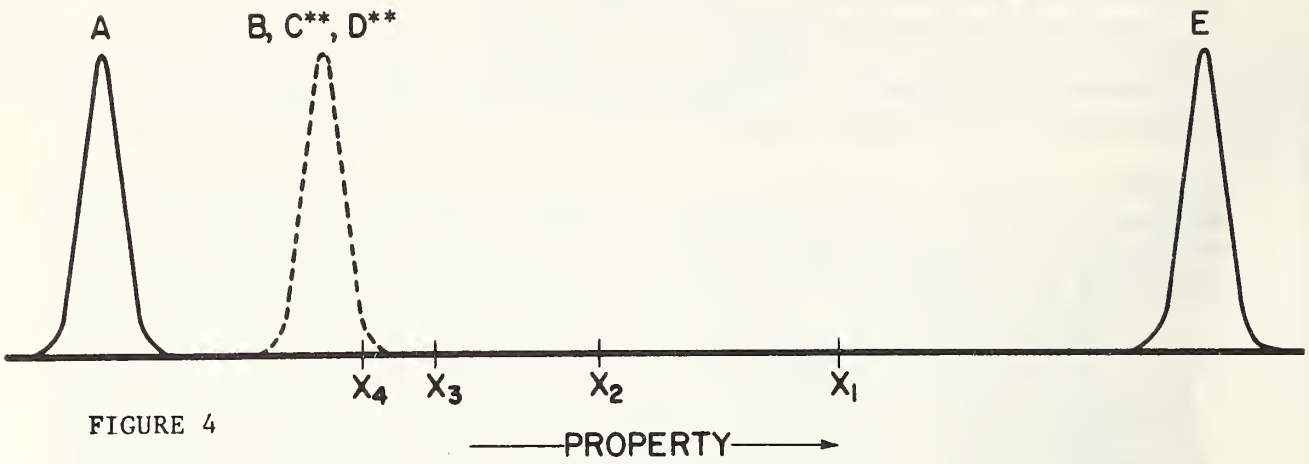


FIGURE 4

Thus, we have four "reasonable" choices--X_1, X_2, X_3, X_4,--at which we can set our level for the standard.

We might decide that X_1, while it takes E off the market, is too low a level of safety even though D will have to take some action to upgrade his quality (see Fig. 2 for the results of this choice).

We might decide that we will institute our standard in stages: X_2, followed in a few years by X_3. C will have, initially, a minor problem. He immediately institutes higher quality control (see Fig. 3), and prepares for the shifting of his total production to the quality of B (see Fig. 4) in a few years. D will have major problems--he must end at B's level in a few years. He might decide first to shift his whole production to B's level, and then at a later time narrow his production range: . Figure 5

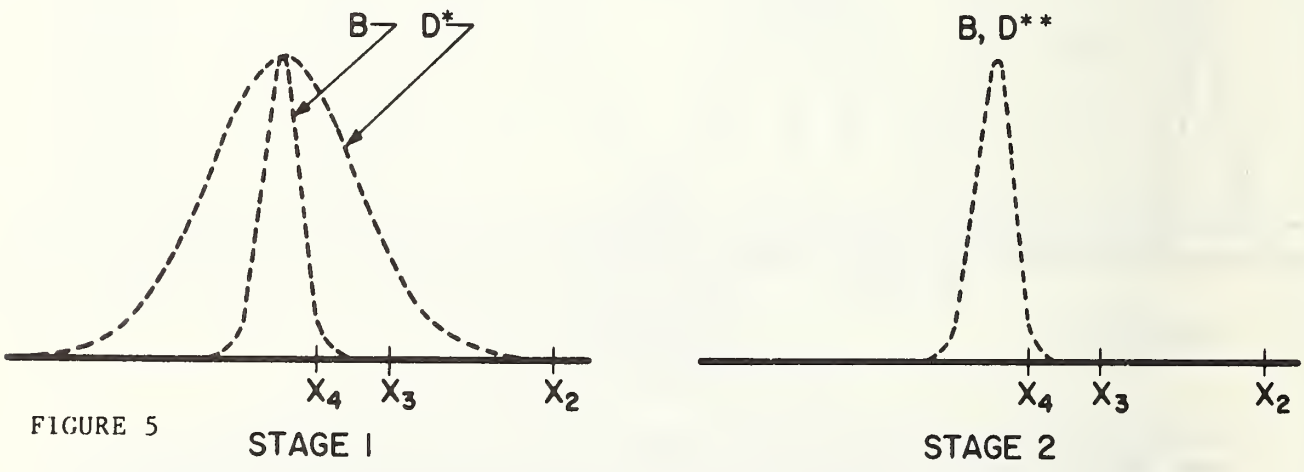


FIGURE 5

or he might first upgrade his quality and quality control to pass the first stage of the standard, and then upgrade his quality to pass the second stage: Figure 6.

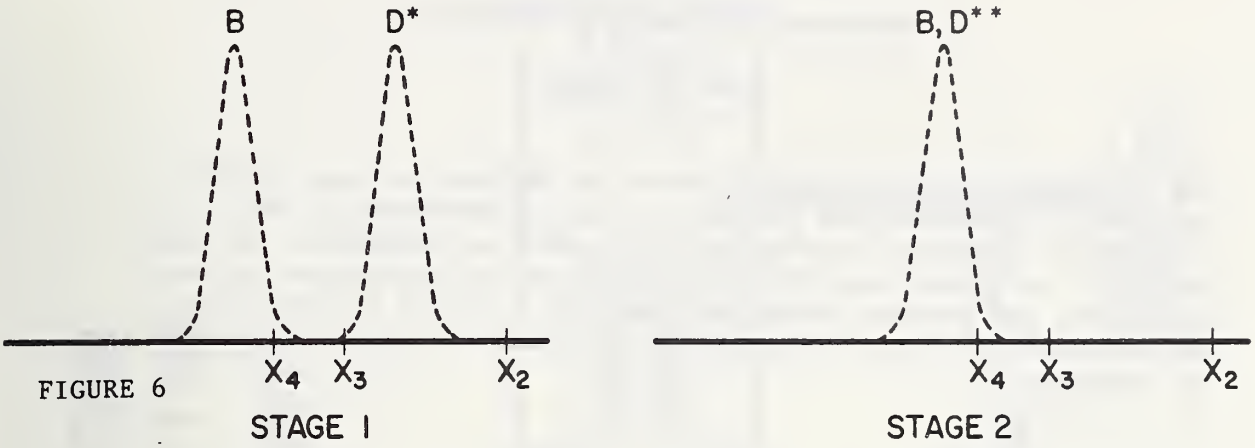


FIGURE 6

Now X_4 certainly provides a higher level of safety than the other three we have just discussed, and we might decide that this is the level we shall set since it is technologically feasible (i.e., all A's pass, most of B will pass, C and D can attain B's level). We would not institute the standard at this level immediately; rather, we would consider it as a final stage reached after manufacturers had first attained intermediate levels X_2 and X_3 at times specified in the standard. In our discussion of the consequences of setting the level at X_3, we noted that both C and D would have problems meeting X_3 even when it was introduced in stages. X_4 is more difficult than X_3 for C and D to attain, and we found at the outset that we did not wish to exclude C and D from the market.

It should be noted that we have shown X_4 in our figures at a level which is not attained by all of B's production. (Neither will it be attained by all of C or D when they reach B's quality level.) We know, *a priori*, that this will be the situation for any level we decide to set in our standard. For example, suppose we set our level at X_3 and insist that all items meet this level. Suppose that the test we use to measure the property associated with the hazard is very difficult to perform, or very expensive, or destroys the product. We cannot test then every item to see that it meets this level; only some of them can be tested. This means we are sampling the production, and there is a chance some items which do not meet the standard will slip through and be sold. We are thus drawn into the area of acceptance sampling, since we wish to know what these chances are.

APPENDIX B

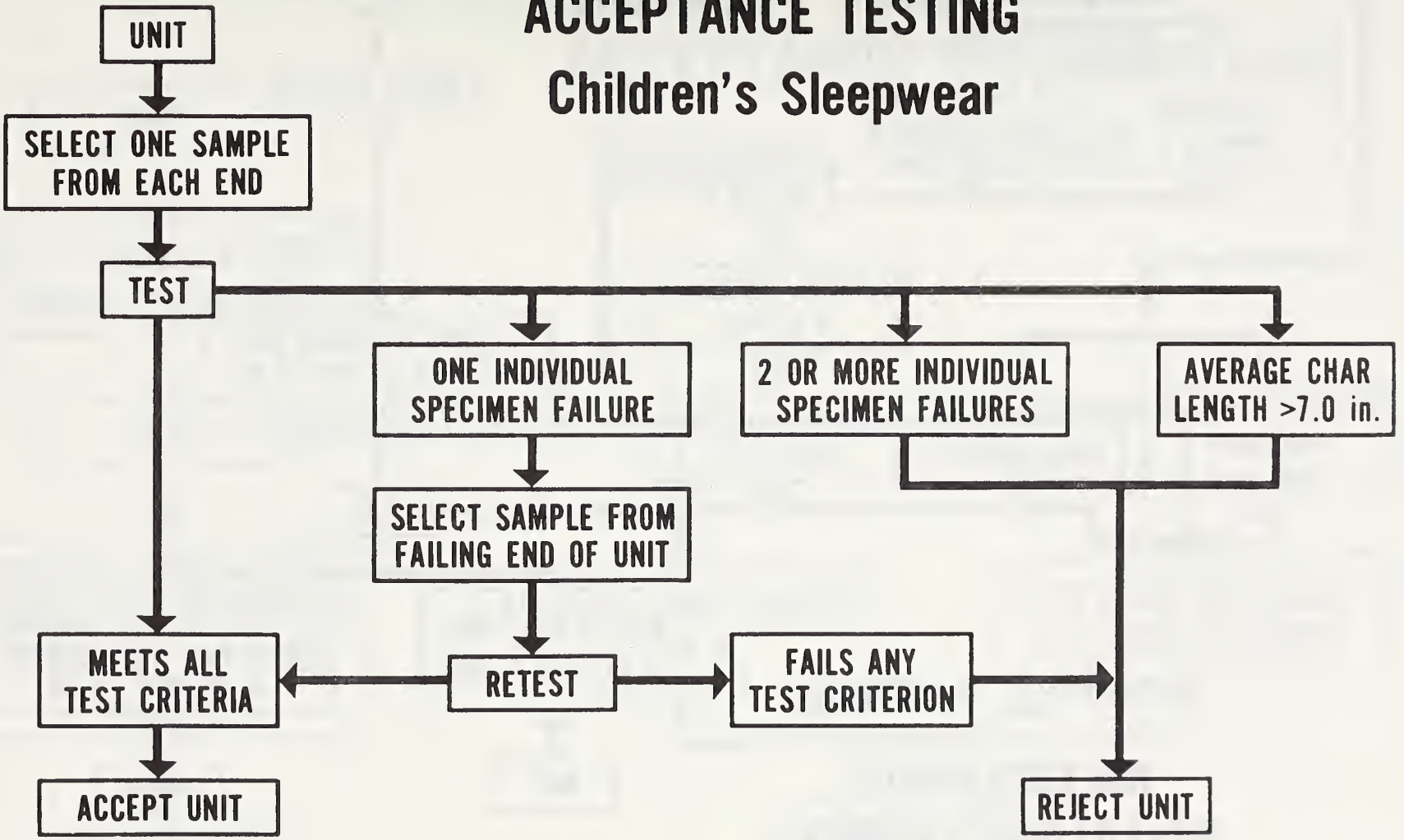
DECISION DIAGRAMS FOR CHILDREN'S SLEEPWEAR FLAMMABILITY STANDARDS

James H. Winger

The following decision diagrams show the decision process which is followed by manufacturers subject to the requirements of the childrens's sleepwear standards to determine acceptability of their product. All three sampling procedures, fabric production, garment prototype, and garment production, must be followed to attain the desired level of fire resistence in the final products purchased by the consumer.

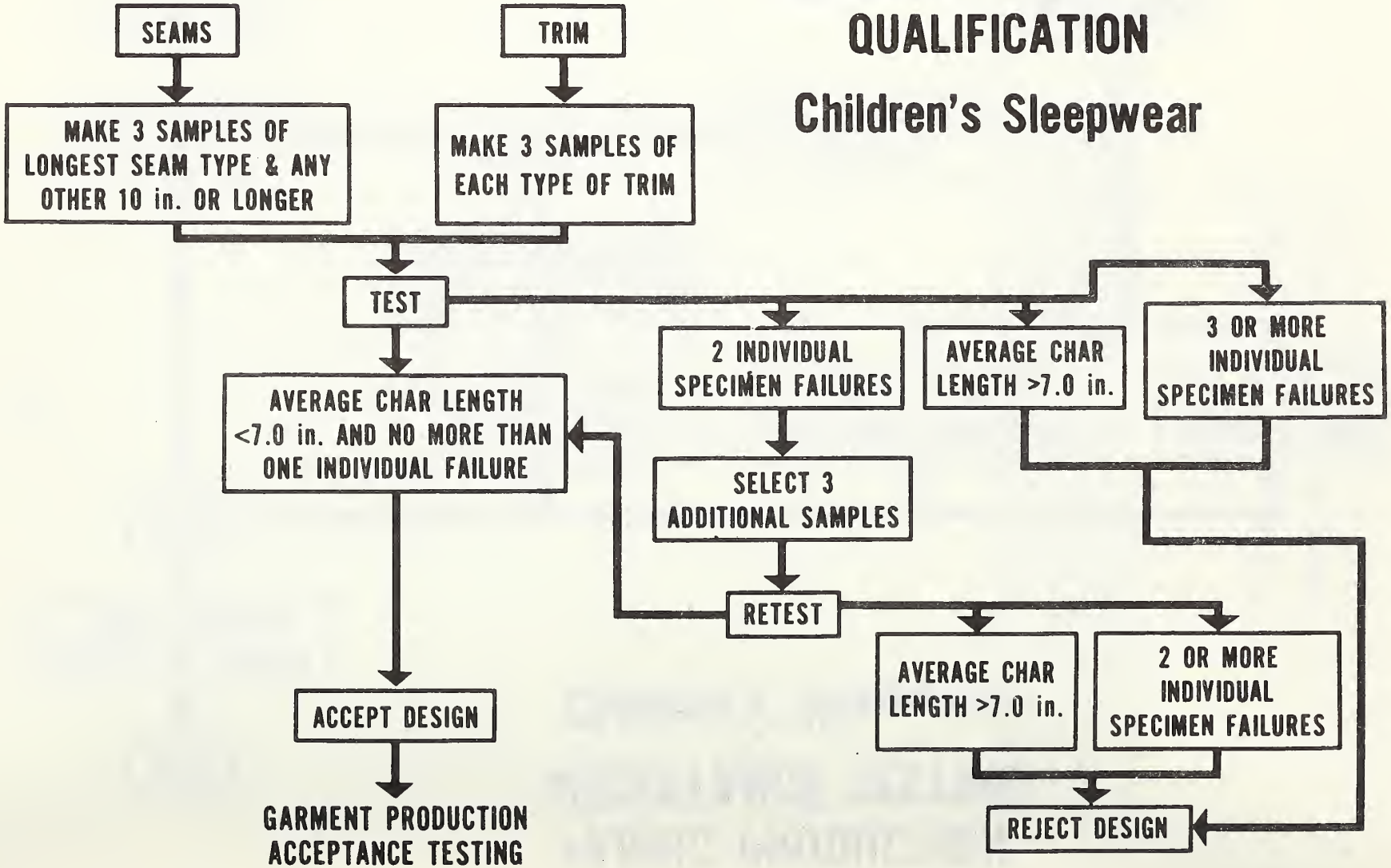
FABRIC PRODUCTION ACCEPTANCE TESTING

Children's Sleepwear



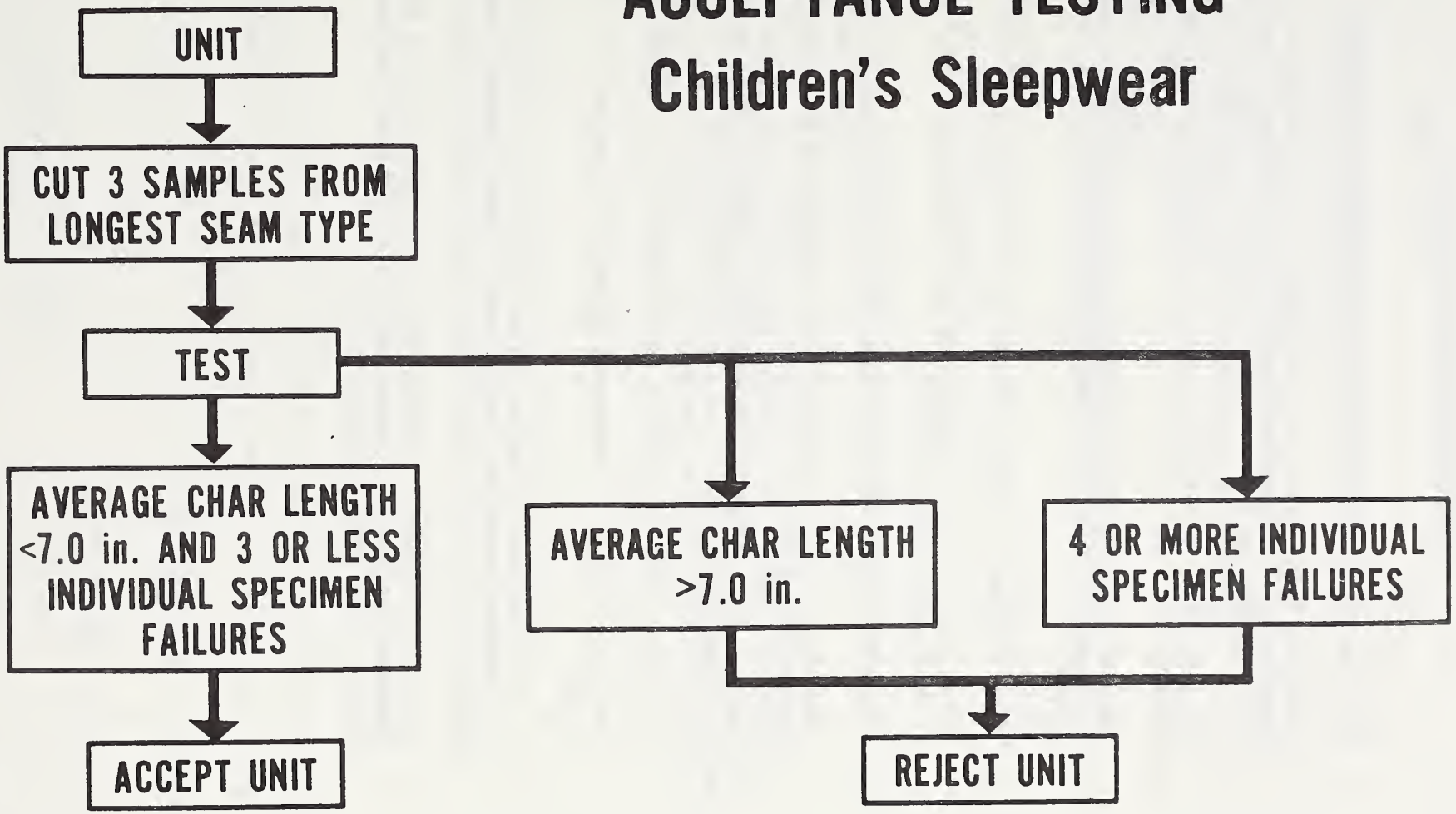
GARMENT PROTOTYPE QUALIFICATION

Children's Sleepwear



GARMENT PRODUCTION ACCEPTANCE TESTING

Children's Sleepwear



Additional Information Sources

1. From the Files of the Consumer Product Safety Commission:

   "Sampling Plans--Need for Development of Applications, to Compliance," information item by R. L. Madison, P. Gottfried, dated 9/8/73

   "Alternatives to Sampling Inspections" by R. L. Madison, dated 9/13/73

   "Recommended Compliance Procedural Philosophy (for Standards Incorporating Sampling Plans) draft by P. Gottfried, dated 10/19/73

   "Some Compliance Sampling Plans" by P. Gottfried, dated 11/12/73

   "Enforcement Without Compliance Market Sampling Plans" by P. Gottfried, dated 11/12/73

2. From ASTM Standardization News, Vol. 2, No. 4, April 1974:

   "Mandatory Standards--Statistical Overview is Necessary" by E. Shecter, p. 8

   "Enforcement of Government Mandatory Product Standards" by A. J. Duncan, p. 12

   "The Evaluation of Standard Test Methods" by J. Mandel, p. 17

   "Manufacturing to a Government Mandatory Standard by J. H. Reynolds, p. 20

3. J. R. Rutherford (Canadian Carpet Institute), "Statistical Aspects Of Consumer Protection Legislation." Paper presented before the American Statistical Association, Ithaca, New York, May 31, 1973.

4. Robert C. Sugarman, "Statistical Problems Associated With Flammability Tests." Report CAL No. QJ-5214-B-1, Calspan Corp., Buffalo, New York, December 8, 1972.

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 75-697	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

CONSIDERATIONS IN THE USE OF SAMPLING PLANS FOR EFFECTING COMPLIANCE WITH MANDATORY SAFETY STANDARDS

5. Publication Date: June 1975

6. Performing Organization Code

7. AUTHOR(S) V. L. Broussalian, A. J. Farrar, J. W. Lyons, J. H. Winger, C. O. Muehlhause, M. G. Natrella, J. R. Rosenblatt, R. D. Stiehler

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Complete Address (Street, City, State, ZIP)

Same as Number 9

13. Type of Report & Period Covered: Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

Various means available to a regulator for gaining compliance with mandatory safety standards are examined. Particular attention is given to his option of mandating a sampling plan or scheme along with the standard. It is concluded that this option as well as the others identified are all viable under suitable conditions and should be available to the regulator for his possible application on a case by case basis.

17. KEY WORDS (six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)

Sampling Plan, Sampling Scheme, Safety Standards, Regulation , Quality Assurance

18. AVAILABILITY

[X] Unlimited

[ ] For Official Distribution. Do Not Release to NTIS

[ ] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13

[X] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT): UNCLASSIFIED

21. NO. OF PAGES: 57

20. SECURITY CLASS (THIS PAGE): UNCLASSIFIED

22. Price: $ 4.25

USCOMM-DC 29042-P74