Pooling and Other Designs for Analyzing Laboratory Samples More Efficiently

by

Walter T. Federer
Cornell University

BU-1036-MA

June, 1989
Revised January 1993
May 1993
August 1993

ABSTRACT

Statistical design procedures exist for retaining anonymity of respondent's answers to sensitive, incriminating, or embarrassing characteristics, for pooling samples in disease prevalence and drug usage studies, and for efficient handling of overloads of analyses in analytical laboratories. They allow results not obtainable by other methods, decrease costs of sample laboratory analyses, allow for timely results, and in general, increase the efficiency and reliability of investigations. Payoffs can be substantial as illustrated with several samples. Sampling procedures and designs allowing estimation of population parameters while retaining anonymity for the respondent are required for many situations in survey sampling. Three such designs accomplishing this are randomized response, block total response, and randomized block total response. Sampling designs which can greatly reduce the number of laboratory analyses for a survey, appear little used and even unknown, e.g., the 1943 Dorfman group testing procedure. This particular group testing procedure as well as other methods of grouping or pooling are discussed to some extent. Three specific surveillance plans are briefly described. Sequential analyses of samples from a survey may be appropriate for some surveys and should be used. The examples and procedures described demonstrate the need for considering and perhaps using results from all areas of statistical design.

1. INTRODUCTION

Since it is essential to know the prevalence of venereal diseases and drug usage in populations, some statistical design assuring anonymity for an individual's response may need to be used. One such design is randomized response, originally proposed by Warner (1965, 1971), and a second is the block total response approach suggested by Raghavarao and Federer (1973, 1979). The Freedom of Information Act does not always protect the rights of the individual having the information but only the rights of the individual seeking the information. If valid information on the prevalence of disease in a dairy herd, say, is to be obtained, then some procedure must be devised that will assure the herd owner that the disease status of the herd does not become public. At present, disease status of a herd when analyses are done by the New York State Diagnostic Laboratory, falls in the public domain. As a result the herd owner can suffer serious economic losses and hence may not cooperate in a survey for disease prevalence. Procedures can be devised to circumvent this, but the herd owner needs to be convinced of this if accurate information is to be obtained (see e.g., Federer *et al.*, 1985).

In many situations, it is desirable to maintain anonymity with regard to an individual's alcohol, disease, or drug status. Some procedures have been and can be devised which will assure anonymity. However, once evidence that a significant proportion of the population is involved in alcohol, drug usage, and/or are infected with a disease, the organization doing the survey will have evidence that a problem exists and can then institute various programs for solving the problem. These items are discussed in Section 2.

A fairly fixed idea among researchers is that individual analyses are required for each individual sample. The idea of pooling samples is foreign to many. If the goals of an investigation can be achieved even when samples are pooled or if the variation among like samples is small, pooling samples may and can be very cost- and time-effective. Examples are presented in Section 3 to illustrate the considerable savings achieved through the technique of pooling samples for laboratory analyses.

In Section 4, several examples are used to illustrate the effectiveness of various pooling procedures. Statistical designs can effectively be used for a variety of situations. In Section 5, a discussion of three specific surveillance designs is given. With the availability of computing facilities, it is shown how to utilize fractional factorials and supersaturated fractional factorials for computer simulations and for screening factors. The use of sequential procedures for analyzing samples from a survey is illustrated with an example in Section 7.

2. ANONYMITY

In surveys involving sensitive, embarrassing, and/or incriminating questions, respondents with the characteristic may be reluctant to cooperate and answer correctly. If respondees are convinced that answers cannot be traced back to them, they are more likely to answer correctly. Note that they *need to be convinced* . It is not sufficient to develop procedures that preserve anonymity of responses but also the respondee must believe that answers are anonymous. As one respondee remarked, "If you people are smart enough to devise these procedures, you are smart enough to find a way to trace my answer back to me." This comment was made on the three survey procedures compared by Smith *et al.* (1974). Also, if respondees feel that class action will occur as a result of their responses, they will answer so this does not happen. Assuming that respondees are convinced of anonymity is insufficient in surveys. They need to be convinced of anonymity and of the reasons for giving correct answers.

One procedure for preserving anonymity is the randomized response (RR) as first proposed by Warner (1965, 1971) and further discussed by Folsom *et al.* (1973), Greenberg *et al.* (1969a, 1969b), and Campbell and Joiner (1973). For one form of RR, the respondee is presented two questions, one sensitive and the other not sensitive, and the respondee answers one of the two questions with known probabilities, say 70% and 30%, respectively. A randomizing device is used to determine for the respondee which question is answered but the question answered is unknown to the interviewer. For example the sensitive question could be, "Do you have HIV?" and the nonsensitive one could be, "Is your random number 7 or 8?". If the randomly selected number is 0,1,...,6, the respondee answers the sensitive question and if it is 7, 8, or 9, the nonsensitive one is answered. Knowing the probabilities that each question will be answered and the probability of answers to the nonsensitive question, it is possible to calculate the proportion of individuals answering yes to the sensitive question in the sample survey. Note that the individuals answering yes to the sensitive question cannot be identified, but only the proportion in the sample answering yes can be obtained.

Raghavarao and Federer (1979) suggest a block total response (BTR) as an alternative to RR. For v questions in the survey, the respondee answers $k \leq v$ of the questions and gives only the sum of the answers to the k questions. Since incomplete block designs were used, the set of k questions is a block and the block total (BT) is the response received from the respondee. Two particular experiment designs which can be used for this situation are a balanced incomplete block design (BIBD) and a supplemented block design (SBD). The latter has quite good variance properties for the sensitive question since it is asked of everyone in the survey. For the remaining v - 1 questions in an SBD, the variance properties will not be as good since these questions are only answered by a fraction of the respondents in the survey.

A modification of the BT procedure of Raghavarao and Federer (1979) was used by Smith *et al.* (1974) in that the particular block of questions answered by a respondent was selected at random by the respondee. This double degree of anonymity was sufficient to assure most respondents of their anonymity.

Another procedure that has been used many times over the years is to use unmarked questionnaires and have these mail-returned in unmarked envelopes. Many respondents feel that there may be some invisible marking or coding on the questionnaire or on the envelope which could expose their identity; hence, they may give non-incriminating and evasive answers to sensitive questions. In the 1940's, Arnold J. King used a locked box on a truck in which responding farmers were requested to deposit their questionnaires. Their answers were written on blank sheets of paper, which could be furnished by the respondee if he felt so inclined. Also, the respondee could observe that there were a large number of other answered questionnaires in the box which would be shaken to mix the questionnaires after he deposited his. Every precaution was taken to assure respondents of their anonymity. Results from this survey on farmers' incomes were completely different from any survey using direct and non-anonymous questioning on farmers' incomes. Correct, or at least more correct, information was obtained because the farmers were assured that the Internal Revenue Service could not identify them. However, if the respondents believed that these results would cause the Internal Revenue Service to audit farmers more frequently, correct answers still would not have been obtained. If the respondees feel that a class action might result if correct answers are given, non-incriminating answers will still be given.

3. POOLING SAMPLES FOR ANALYSES

In many surveys, a part or all of the responses come from analyses done in an analytical laboratory. For example, in a large dietary-cancer mortality survey conducted jointly by Beijing and Cornell Universities, biochemical analyses were obtained for approximately 300 characteristics. If analytical analyses were conducted on individual blood samples, the quantity of blood from one individual is insufficient for the 300 analyses. Furthermore, if each sample was analyzed individually, the laboratory facilities would be strained far beyond capacity. In the planning stages of the survey, the author pointed out that the researchers needed to reconsider their goals, to note that cancer mortality came by sex and by county, and to consider pooling the samples for the 25 males and the 25 females in each of the two communes in a county. This was to be done for each biochemical analysis for which pooling did not cause interactive effects; analyses for hemoglobin, e.g., would not be obtained from pooled samples. This would result in doing only 4% of the analyses originally proposed. In addition, a large enough pool of blood

would be available to perform all the necessary analyses. The suggestion was adopted, and a part of each sample was frozen and stored for possible future analyses (see Federer *et al.*, 1986)

In performing laboratory analyses for scientific studies, the failure to report how quality and accuracy were maintained is a serious deficiency. Many researchers do not appear to realize the necessity for using quality control procedures and being certain that standards and controls are used consistently. Several such cases have been encountered during the course of statistical consulting. Statisticians need to develop quality control procedures for laboratory analyses and to become actively involved in their application and use in practice. In the Beijing-Cornell dietary-cancer mortality study, it was recommended that ten check samples at a low level, at a medium level, and at a high level of the characteristic being analyzed, be performed. The 30 check samples were to be randomly distributed in with the 2(2)(65) = 260 pooled samples for each biochemical analysis conducted. These check sample results can be used to obtain an estimate of error variation for the particular analysis and analyst under consideration. Reporting the results for the check samples along with the study results would lend credibility to the investigation.

From consulting experience, it appears that the lack of quality control and of controls on analysts and procedures in laboratories requires considerable attention from the statistician. The procedures used are often highly inefficient as well as lacking adequate controls. The net result is that when a scientific paper is published, it usually is impossible for the reader to ascertain if the original data are reliable or unreliable. It should be a requirement for papers published in scientific journals to state clearly and precisely the quality and accuracy of the procedures used.

A technique known as group testing is attributable to Dorfman (1943). The procedure was supposed to have been developed for making possible the very large number of tests for venereal diseases required by the United States Defense Forces in World War II. It is stated in Feller (1950) that use of the procedure by the U.S. Army resulted in up to 80% savings. Group testing should be common practice in analytical laboratories. Pooling procedures have, however, been used by some investigators (see e.g., Comstock *et al.*, 1987; Mortimer, 1980; Wahrendorf *et al.*, 1986). As with the dietary-cancer mortality example which required only 4% of the originally proposed analyses, group testing can result in considerable savings in analyses (see e.g., Federer, 1983, 1991). The Dorfman procedure is useful for all laboratory analyses for which the following conditions hold:

(i) The presence or absence of a characteristic.
(ii) The percentage of positives is relatively small.
(iii) The composite sample mean is the same as the mean of the individual samples.
(iv) The compositing does not alter the characteristics of the samples.
(v) The laboratory technique is accurate enough to detect 1/g of the substance for g the group size used.

When the above conditions are satisfied, group testing can save a considerable amount of time, laboratory equipment, and personnel. Since such savings can be substantial and since laboratory directors and analysts appear to be unaware of the procedure, effort needs to be made to acquaint them with the procedure. Some new procedures will need to be devised by statisticians in order to meet the needs of investigators.

For N samples to be screened for a characteristic and a group size of g, there will be $N/g = G_1$ groups that require an analysis for the first phase to determine if the characteristic is present (positive) or absent (negative) in the pooled sample. Let us assume that presence and absence of the characteristic is a binomially distributed variable with population parameter p, the percentage of the population possessing the characteristic. An expected number $G_2 \leq G_1$ groups will test positive for the characteristic. Then for the G_2 groups, the Dorfman procedure would test for the presence of the characteristic by doing analyses on the individual samples of the group. The total expected number of analyses then would be:

$$G_1 + gG_2 = N(1/g + 1 - q^g), \tag{3.1}$$

where $q = 1 - p$. When

$$1/g = q^g \text{ or } q = (1/g)^{1/g}, \tag{3.2}$$

it will be a toss-up as to whether to do individual samples or to use group testing. This is called the break-even point, which occurs when p is approximately 0.29.

When p is small, considerable savings can be achieved by using the optimum group size which is obtained from an iterative solution of:

$$g^2 = (q^g \log (1/q))^{-1} \tag{3.3}$$

For $p = 0.05$ and $q = 0.95$, the optimum group size is 5 and the expected total number of analyses required is 0.426N, or a savings of 57.4%. For $p = 0.01$ and $q = 0.99$, the optimum group size is 11 and the expected total number of analyses required is 0.196N, or a savings of 80.4%. As p becomes smaller, the savings increase. Also, as p becomes smaller, variation in group sizes around the optimum has smaller effect on the savings.

Some recent work on Dorfman's procedures as well as on a modification has been reported by Huang *et al.* (1989) and Johnson *et al.* (1989). They discuss the effect of faulty inspection on the Dorfman procedure and on several modifications of the method. Since analytical tests are not always perfect, i.e., a characteristic may be present but it goes undetected in a particular analysis, it is desirable to know the properties of a procedure in the presence of possible faulty inspection. The authors have provided answers to these problems, and their work may be considered to be a first step in setting up quality control procedures under group testing.

As has been demonstrated above and in Federer (1984, 1991), there are many methods of pooling beside group testing. The statistician and the investigator will need to use creativity in developing new pooling procedures to meet the needs of investigators. Another such procedure, with variations, is described below.

It is sometimes possible to have a very accurate test available for detecting a drug or a disease and a pool of many individuals can be used. For example, it has been demonstrated that equal amounts of 75 blood samples from cows can be pooled and the ELISA test is sensitive enough to detect the presence of *Bovine leucosis* virus (BLV) if one cow in the 75 has BLV (see Federer *et al.*, 1985). When such sensitive procedures are available and when the researcher knows the mean titer levels of diseased and disease-free cows in the population (or drug and no-drug levels), one laboratory analysis on the pool is sufficient to estimate the approximate number of individuals having the disease. To illustrate for a disease like BL and a virus like BLV, suppose the mean titer level for disease-free animals is M_f (will often be zero for a characteristic), the mean titer level for diseased animals is M_d, the titer level of the pooled sample is M_s, and the number of animals in the pool is N. Let N_d be the number of diseased animals and $N - N_d$ be the number of disease-free animals. Then,

$$N_d M_d + (N - N_d) M_f = N M_s, \tag{3.4}$$

and

$$N_d = N(M_s - M_f)/(M_d - M_f). \tag{3.5}$$

When M_f is zero,

$$N_d = N M_s / M_d. \tag{3.6}$$

The prevalence of the disease is N_d/N.

In testing samples for alcohol and drugs, say, M_f will be zero. The value for M_d may be obtained by summarizing the many tests that have been made previously. The distribution of alcohol levels in blood samples of drivers picked up for drunk driving is available in abundance; M_d then is obtained as the arithmetic average of these tests. From the distribution of levels, procedures need to be devised for obtaining confidence levels for the point estimates N_d and N_d/N. Likewise, drug levels in urine samples of drug users are available; here again a value for M_d is obtainable. Actual titer levels in cows are obtainable, but often not recorded, and obtaining values for M_f and M_d is only a bookkeeping problem. If all that is required by an organization is a "quick and perhaps dirty" estimate of the prevalence of the problem, then use of the above idea can cut laboratory costs considerably and can maintain anonymity.

To illustrate how some of the ideas may be used (see Federer, 1984 and 1991, for other ideas on pooling), let us consider drug testing in urine samples for a given organization such as, e.g., a company, a school, an Army unit, etc.. To preserve anonymity and to use pooling, let us consider the following situation for an Army unit. When all members are on

the base and since they all are regimented to arise at the same time, one of the first things they do each morning is to urinate. If the urinals are built so that all the urine for one hour, say, can be collected in large bottles, all the urine for that period of time could be collected. The bottles would be small enough so that if one person were using drugs, it would be detectable. An automatic or manual shutoff could be installed so that when a bottle was filled, no more urine ran into the bottle. One sample would be taken from each bottle and analyzed for the amount of drug present in bottle b, say $M_{sb}=N_{db}M_d/N_b$. Knowing the approximate total amount of urine collected and the total number of individuals urinating, the number of individuals urinating in bottle b, N_b, could be approximated. Then from formula (3.6) for M_f equal to zero, the number, or proportion, of individuals using the drug could be estimated, provided M_d is known. The estimates for individual bottles could be averaged to obtain an estimate for the population under consideration. Knowing the approximate prevalence of drug usage, the commander could, e.g., decide to instigate an anti-drug usage campaign, to continue testing over time, to do random testing of individuals, or to test all individuals. If drug usage decreased to within acceptable limits from an anti-drug usage campaign, e.g., individual or random testing of individuals may not be instituted. If so, anonymity would have been maintained for all individuals with respect to previous drug usage.

4. GROUP TESTING AND DESIGN FOR DRUGS AND DISEASES

A form of Dorfman's (1943) group testing would have been very cost-effective and time-effective for the following example encountered by the New York State Diagnostic Laboratory (NYSDL). For the particular breed of horses involved, breeding must be done by natural rather than artificial means in order to have the foals registered. It was suspected that some mares might be infected with a newly found venereal disease. None of the breeders would use their stud on the mares until they had been certified free of the disease, as an infected stud would become useless. Blood samples from 600 mares were sent to the New York State Diagnostic Laboratory for analysis for the disease. The NYSDL charged $100 per sample for analysis, and the job required NYSDL's facilities for a week. Now,if group testing had been used, 30 say, samples could have been combined to form a pool, provided the test for the disease was sensitive enough to detect one diseased mare in a pool of 30. Then for those pools testing positive, pools of ten would have been made; from the positive groups, pools of five would have been made. For any positive pool of five, individual samples would then be tested. If no mares were infected, only 20 analyses would have been made, resulting in a fee for each analysis of $60,000 / 20 = $3,000, considerably more profitable to NYSDL than $100 per analysis. If one mare had been infected, 20 + 3 + 2 + 5 = 30 analyses would have been made, resulting in $2,000 per

analysis. The prevalence of the disease was suspected to be small. If the sensitivity of the test had been very high, a more efficient procedure (see Sobel, 1967, and Sobel and Groll, 1959, 1966) would have been to test a pool of 600 first; if positive, split into pools of 300. For positive pools of 300, split into pools of 150; for positive pools, split into pools of 75. Then for the positive pools of 75, split into pools of 37 and 38; for positive pools, split into pools of 18 or 19, then pools of 9 or 10, then pools of 4 or 5, then into pools of 2 or 3, and finally into individual sample analyses. For one infected mare, 19 or 20 analyses would have been required; for two infected mares, at most 39 analyses would have been required. This procedure requires that the quantity of blood available be sufficient for that many splittings. Other procedures are available (see Raghavarao and Federer, 1973, and Bush *et al.*, 1980).

Pooling procedures described herein can be very cost-effective and time-effective, as well as preserving anonymity of an individual's status. Collecting individual urine samples would be time-consuming and costly. The pressure for so many individuals to urinate in a short period of time as for the Army example, would make the task almost impossible without a large work force or else the individuals would have to be required to urinate in a bottle with their identification on it. In any event, this would be an organizational task of considerable dimension.

If tests for several diseases and/or drugs are required, a sample from one individual may be too small to perform all analyses. For example, blood sample tests for 5 to 10 different diseases and 10 to 20 different drugs may be desired. Blood from one individual may be insufficient for this. By using pools, sufficient blood would be available, as was the case for the dietary-cancer mortality study.

If anonymity is required as well as having a larger sample than can be obtained from one individual, the following BIBD procedure is suggested via an example. Suppose $v = 7$ individuals are to be tested and if pools of $k = 3$ are needed for analysis, the following procedure, assuring anonymity and a variation of one given by Smith *et al.* (1974), is suggested. Have an individual select at random and anonymously a letter from A, B, C, D, E, F, and G, without replacement. The individual is instructed to divide his urine or blood sample into three equal parts and place one part in each of the three pools where his letter occurs as follows:

Pools						
1	2	3	4	5	6	7
A	B	C	D	E	F	G
B	C	D	E	F	G	A
D	E	F	G	A	B	C

For example, an individual selecting letter B would place his sample in pools 1, 2, and 6. The remaining individuals do likewise. The pools should be arranged so that the technician cannot see which pools the individual is using. If the sample is urine, an individual could pour his collected urine into the three containers and then deposit them in the appropriate pools. If the sample is blood, the technician taking the sample could partition the sample into three parts and the individual could then deposit them in the appropriate pools. The samples from all individuals should be of the same size. From the pool totals, it is possible to obtain solutions for each letter. Note that if the pools and/or the identity of the letter are known, anonymity is not maintained. BIBDs for $v = b$, $k = r$, and a pairwise balance for all v items in the b pools (blocks) occur for many values of v and k (see e.g., Federer, 1955, Table XI-1; Raghavarao, 1971, Table 5.10.1).

5. SURVEILLANCE PROCEDURES

Surveillance procedures for diseases and drugs are a part of survey design. Although many surveys are usually one-shot affairs, the idea of continuous sampling over time to determine the status of a population is not new. The Bureau of Labor Statistics has an ongoing survey to do surveillance of employment status of the United States population. During election campaigns of political candidates, the status of a candidate's chances of winning is under constant scrutiny. The same type of surveillance is used for many diseases. With respect to a disease which is to be eradicated or to bring to a very low prevalence, some different problems arise than for the surveillance designs of the above nature. To illustrate, let us consider the disease *Bovine leucosis* (see Federer *et al.*, 1985), where it is desired to eradicate the disease in individual dairy herds. The present prevalence is of the order of 30% to 50%, or possibly even higher. Simple management procedures such as disease-free semen, clean needles, clean dehorning devices, and giving first priority to culling diseased cows, can be used effectively to reduce prevalence. Economic costs of these procedures are low. The first problem is to identify the cows having BLV, which has to be done without using pools because of the high prevalence. Once the prevalence is under, say, 20%, group testing can be used effectively in screening samples for BLV. It would be desirable to use milk or urine samples rather than blood samples because the last needs to be done by a professional whereas the former could be collected by the dairyman. It is not known if cows can be screened for BLV using milk or urine samples at the present time. When the prevalence is reduced nearly to zero, questions of optimum samplings through time and size of pool arise. Once the prevalence of BLV has been estimated as zero, it may be necessary to sample the herd only once every two to three years to ascertain the disease-free status of the herd. Optimum sampling intervals need to be resolved for each case. Perhaps one pool for the entire herd may be all that is

necessary, provided that the test for BLV is that accurate. Note that the BLV problem is much like the HIV drug needle situation.

The control and surveillance procedures for another disease of dairy cattle, Blue Tongue, are very different from BLV (see Clark *et al.*, 1985). Blue Tongue is a viral, insect-borne infectious disease of domestic and wild ruminants as well as dairy cows. Instead of working on a herd basis as with BLV, it is necessary to work on a regional basis to account for the mobility of the insect carrying the virus. Procedures for doing this are described in Clark *et al.* (1985).

As a third example, the NEW York State Health Department desires to have procedures to determine whether or not cases of leukemia are occurring in clusters over time. Once it has been established that a cluster of leukemia cases exists, an investigation could be initiated to determine possible causes such as contaminated air, water, food, or other causes. Presently, about seven or eight years of data exist relative to the time and geographical location of each new case of leukemia. As more and more data are collected, it is desired to know if and where clusters are occurring. One surveillance procedure suggested to date is to determine the probability that a cluster exists at a given time. If this probability is, say, 75%, this cluster should be kept under close scrutiny. If the probability that a cluster exists increases to 90%, say, an investigation into probable causes would be initiated. Improvements in clustering procedures are needed as well as a study of what type of data to collect.

6. COMPUTER SIMULATIONS AND FACTOR SCREENING

As pointed out by Federer (1987a, 1987b) and references therein, statistical design can be used effectively to greatly reduce the number of combinations for experimentation via simulations. Statisticians use the computer for their experimentation using simulations, specific examples, limits, etc. Since many factors are often involved, the total number of combinations is usually large. Fractional replication is a very useful statistical design for many situations. To illustrate, Grimes and Federer (1984) were interested in simulations involving four levels each of significance, of variance inequality, of sample size configurations, and of parameter structure. The total number of combinations is $4^4 = 256$. It was decided that a main effect fractional replicate would suffice for determining the effects of these four factors. An orthogonal array of 16 combinations was used as the fractional replicate plan, resulting in solutions for all main effects and leaving three degrees of freedom for lack of fit. This resulted in running simulations using only a 1/16 fraction of the complete factorial and the desired results were obtained. They also used the computer to aid in a numerical integration problem. Other similar examples have been encountered in the course of statistical consulting. Investigators should make use of

fractional factorials in conducting computer experiments involving simulations, but many appear to be unaware of this possibility. Main effect plans should be especially useful but Resolution IV, V, and perhaps higher (see Raktoe *et al.*, 1981), may be needed in certain situations.

In addition to the above mentioned fractional factorial designs, there are supersaturated fractional replicates which group levels of several factors together. Steps for using and setting up these statistical designs are discussed in Federer (1987b). It is unfortunate that the term group testing has also been applied to supersaturated designs (see Watson, 1961, and some of the references in Federer, 1987b). The term should be reserved for the situation envisaged by Dorfman (1943). Such designs, if appropriate, can greatly reduce the number of combinations at which to run experiments and can be very useful in reducing the number of factors and factor groups. Factors can be screened through use of supersaturated designs.

In investigating characteristics of certain diseases or drugs, it may be possible to use fractional factorials and supersaturated fractional factorials. In establishing the accuracy and sensitivity of disease and drug tests, several factors at several levels may need to be investigated. Certainly, the above designs would be useful.

7. SOME OTHER SITUATIONS

Federer (1984, 1991) describes a number of situations where group testing and other statistical designs can be and should be used. One such procedure is sequential sampling of samples from a survey. To illustrate, consider the Love Canal chemical dump situation of a few years ago. It was imperative that all samples be taken quickly and as extensively as possible to determine contamination, if any, at various distances from the chemical dump. Granted that all the samples had to be taken because of political pressures but not all samples had to be analyzed and certainly not individually. If a sequential procedure for analyzing the samples had been used, considerable monetary and other savings would have accrued. The cost for analyzing each sample was in the $2,000 range. An efficient sequential method would have been to analyze the samples, or a pool of samples, near the chemical dump first with sparse sampling of areas further away from the dump. If there was no contamination, no further samples would need to be analyzed. This project reportedly cost around 4.5 million dollars. The suggested sequential procedure could have saved a sizeable portion of the 4.5 million.

Sequential sampling plans for spread of a disease from a focal point should be utilized wherever possible. Likewise, survey samples most likely to show an effect should be analyzed first. If no effect is observed, it may be possible to forego analyzing the remaining samples. Such schemes will result in considerable savings.

Another form of pooling may be possible for certain types of surveys. Suppose that variation among the smallest sampling unit is very small, e.g., iodine content of soybean samples. In a survey of soybean iodine levels for a State, the samples from counties by variety, say, could be pooled and only one analysis for iodine content on the pool could be conducted. In a designed experiment, a suggested procedure was to pool the r replicates for each soybean treatment and to make one determination for the pooled sample. This resulted in doing 1/rth of the determinations as compared to processing individual samples. Likewise, variation among determinations may be small but variation among samples within a primary sampling unit may be relatively large. An efficient procedure would be to make only one determination on the pool of samples for each primary sampling unit. For example, let the variance among determinations be 1 and the variation among samples be 10. For d = 2 determinations per sample and for s = 2 samples per primary sampling unit (psu) in a survey, the variance of a psu mean is 21/4; for d = 1 and s = 4, the variance of a psu mean is 11/4, or roughly one-half of the previous case.

8. SUMMARY

Many examples have been presented demonstrating the usefulness of statistical design for efficiently and effectively conducting investigations. The broad scope of statistical design covering survey design, surveillance design, sequential design, sampling design, treatment design, allocation of resources, and measurement design is illustrated with examples. Limited outlook on what constitutes statistical design can stifle use and effective investigation. The importance of planning efficient and effective investigations cannot be over emphasized.

Pooling procedures do not appear to be widely used. Considerable savings in costs, personnel, and facilities are possible when appropriate pooling procedures are utilized. Pooling methods take on many forms such as for example, RR and BTR methods, fractional replication (effects are pooled), supersaturated fractions, the Dorfman group testing method, and others. More emphasis in introductory statistics classes (see Federer, 1991), on the topics discussed herein would go a long way toward acquainting investigators with efficient ways to conduct investigations.

9. ACKNOWLEDGEMENT

The comments of a referee were helpful in preparing the final version of this paper. The author wishes to thank the referee for suggestions to improve and focus the manuscript.

10. LITERATURE CITED

Bush, K. A., W. T. Federer, H. Pesotan, and D. Raghavarao (1984). New combinatorial designs and their applications to group testing. *J. Statistical Planning and Inference* **10**, 335-343.

Campbell, C. and B. L. Joiner (1973). How to get the answer without being sure you asked the question. *The American Statistician* **27(5)**, 229-231.

Clark, L. C., E. J. Dubovi, W. T. Federer, J. S. Krantz, D. S. Robson, and A. Torres (1985). A surveillance and control program for Blue Tongue. BU-879-M in the Technical Report Series of the Biometrics Unit, Cornell University, June.

Comstock, G. W. *et al.* (1987). Re: vitamin measurements in pooled blood samples. *American J. of Epidemiology* **124**, 169-170.

Dorfman, B. (1943). The detection of defective members of large populations. *Annals of Mathematical Statistics* **214**, 436-440.

Federer, W. T. (1955). *Experimental Design -Theory and Application.* Macmillan Co., New York (Republished by Oxford & IBH Publishing Co., Calcutta, New Delhi, in 1967 and 1974).

Federer, W. T. (1983). Group testing: Break-even percentage and optimal group size. BU-24-P in the Problem Series of the Biometrics Unit, Cornell University, November.

Federer, W. T. (1984). Cutting edges in biometry. *Biometrics* **40**, 827-839.

Federer, W. T. (1987a). Fractional replication in simulation studies. *Communications In Statistics, Simulation and Computation* **16(1)**, 233-237.

Federer, W. T. (1987b). On screening samples in the laboratory and factors in factorial investigations. *Communications In Statistics, Theory and Methods* **16(10)**, 3033-3049.

Federer, W. T. (1991). *Statistics and Society - Data Collection and Interpretation.* Marcel Dekker, Inc., New York. Chapters 5 and 6.

Federer, W. T., L. C. Clark, E. J. Dubovi, and A. Torres (1985). A surveillance and control procedure for *Bovine leucosis* and other diseases. BU-877-M in the Technical Report Series of the Biometrics Unit, Cornell University, September.

Federer, W. T., L. C. Clark, N. J. Miles-McDermott, and D. S. Robson (1986). Notes on statistical analyses of the data from the Cornell-China Project. BU-917-M in the Technical Report Series of the Biometrics Unit, Cornell University, November.

Feller, W. (1950). *Probability Theory and Its Applications*. John Wiley & Sons, Inc., New York, pp. 189.

Folsom, R. E., B. G. Greenberg, D. G. Horvitz, and J. R. Abernathy (1973). Two alternative questions randomized response model for human surveys. *J. American Statistical Assoc.* **68**, 525-530.

Greenberg, B. G., J. R. Abernathy, and D. G. Horvitz (1969a). Application of the randomized response technique in obtaining quantitative data. *Proc., Social Statistics Section, American Statistical Assoc.*, pp. 40-43.

Greenberg, B. G., A. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz (1969b). The unrelated question randomized response model. Theoretical framework. *J. American Statistical Assoc.* **64**, 520-539.

Grimes, B. A. and W. T. Federer (1984). Comparison of means from populations with unequal variances. In *W. G. Cochran's Impact on Statistics* (editors: P. R. S. Rao and J. Sedransk), John Wiley & Sons, Inc., New York, pp. 353-374.

Huang, Q., N. L. Johnson, and S. Kotz (1989). Modified Dorfman - Sterrett screening (group testing) procedures and the effects of faulty inspection. *Communications In Statistics; Theory and Methods* **18**, 1485-1495.

Johnson, N. L., S. Kotz, and R. N. Rodriguez (1989). Dorfman - Sterrett screening (group testing) and the effects of faulty inspection. *Communications In Statistics; Theory and Methods* **18**, 1469-1484.

Mortimer, J. Y. (1988). Saving tests by pooling sera - how great are the benefits? *J. Clinical Pathology* **33**, 1120-1121.

Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments*. John Wiley & Sons, Inc., New York.

Raghavarao, D. and W. T. Federer (1973). Group testing--A combinatorial approach. BU-473-M in the Technical Report Series of the Biometrics Unit, Cornell University, July.

Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *J. Royal Statistical Soc., Series B*, **41**, 40-45.

Raktoe, B. L., A. Hedayat. and W. T. Federer (1981). *Factorial Designs*. John Wiley & Sons, New York.

Smith, L. L., W. T. Federer, and D. Raghavarao (1974). A comparison of three techniques for eliciting truthful answers to sensitive questions. *Proc., Social Science Statistics Section, American Statistical Assoc.*, pp. 447-452.

Sobel, M. (1967). Optimum group testing. *Proc., Colloquium on Information Theory,* Bolyai Mathematical Soc., Debrecen, Hungary, pp. 411-488.

Sobel, M. and P. A. Groll (1959). Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical J.* **38**, 1179-1252.

Sobel, M. and P. A. Groll (1966). Binomial group testing with an unknown proportion of defectives. *Technometrics* **8**, 631-656.

Wahrendorf, J., *et al.* (1986). Vitamin measurements in pooled blood samples. *American J. of Epidemiology* **123**, 544-550.

Warner, S. L. (1965). Randomized response. A survey technique for eliminating evasive answer bias. *J. American Statistical Assoc.* **60,** 63-69.

Warner, S. L. (1971). A linear randomized response method. *J. American Statistical Assoc.* **66**, 884-888.

Watson, G. S. (1961). A study of the group screening method. *Technometrics* **3**, 371-388.