63-4-2

ORC 63-1-(RR)
30 January 1963

SOME STATISTICAL PROBLEMS OF DIAGNOSIS BY MEANS OF A MULTIPHASIC SCREENING EXAMINATION

by

Ellen Sherman

OPERATIONS RESEARCH CENTER

INSTITUTE OF ENGINEERING RESEARCH

UNIVERSITY OF CALIFORNIA - BERKELEY

SOME STATISTICAL PROBLEMS OF DIAGNOSIS BY MEANS OF A MULTIPHASIC SCREENING EXAMINATION

by

Ellen Sherman
Operations Research Center
University of California, Berkeley

30 January 1963

ORC Report No. 63-1

This research was primarily supported by the Office of Naval Research, Contract Nonr-3656(02) Project NR 043 275 and partially supported by the National Institute of Health Grant RG - 9606 and the National Science Foundation Grant NSF-G 21034 with the University of California.

SOME STATISTICAL PROBLEMS OF DIAGNOSIS BY MEANS OF A MULTIPHASIC SCREENING EXAMINATION

This research report is written at an early stage in the revision and expansion of the multiphasic physical examination of the Permanente Medical Group. The concept of multiphasic screening in medicine refers to a series of tests performed to determine whether there is sufficient likelihood of certain diseases being present to warrant further testing for these diseases. Multiphasic screening examinations must be suitable for routine application to a large number of patients, as well as comprehensive, accurate, and efficient.

The problem is to determine the type of data to be obtained from each patient and a process by which a diagnostic decision is to be reached which will be suitable for handling by an electronic computer. This report is intended to outline a statistical formulation of the problem, to provide references to previous studies in related areas, to point out several statistical problems to be anticipated at various stages in the development of such a program, and to summarize a portion of the ideas set forth by the committee in charge.

The approach which is suggested in this paper involves determining a set of diseases and/or disease classes and for each of these classes a set of diagnostic questions and tests, each having a positive or a negative response. On the basis of a patient's responses to these tests he will be either dismissed as free from these diseases or referred for further diagnostic study in one or more of the disease classes. The number of classes must be large enough to render the examination worthwhile, yet it must be subject to the limitations that cost and convenience place upon the type and upon the total number of questions and tests that are feasible for a multiphasic examination.

The problem becomes statistical in as much as few signs or symptoms, if any, have the property of always occurring when a particular disease is present and never occurring when it is not. In view of the basic approach to the examination as outlined above, it seems possible to consider a sequence of two-decision problems, or problems in testing one hypothesis against another. Suppose, then, there is to be established a set of m disease classes, $C_1,\ldots,C_m$, with a set of simple hypotheses versus simple alternatives, H_i : Patient should be examined more thoroughly with regard to disease class C_i, versus K_i : Patient does not show signs of a disease in class C_i, $i = 1,2,\ldots,m$. Hereafter, the notation shall be $H_i : C_i$, vs. $K_i : C_i^{-1}$.

For each of the m hypotheses, a set of diagnostic questions and tests must be determined to yield a diagnostic vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i})$. The X_{ij} can assume the values zero or one according as the corresponding symptom is absent or present. Hence there are 2^{n_i} theoretically possible points X_i. It is assumed that the probability of observing a point x_i for a patient in C_i^{-1} is different from that for a patient having a disease in class C_i; i.e., $Pr(x_i|C_i) \neq Pr(x_i|C_i^{-1})$, with the possible exception of a few x_i. Hence, we are able to determine a best critical region $\mathcal{C}_i$ for the rejection of the hypothesis H_i, using the Neyman-Pearson fundamental lemma, and to determine the probabilities of error. The probability of rejecting H_i when it is in fact true is given by $\alpha_i = Pr(X_i \in \mathcal{C}_i|C_i) = \sum_{x_i \in \mathcal{C}_i} Pr(x_i|C_i)$, and the probability of accepting H_i when it is in fact false is $(1 - \beta_i) = Pr(X_i \notin \mathcal{C}_i|C_i^{-1}) = 1 - \sum_{x_i \in \mathcal{C}_i} Pr(x_i|C_i^{-1})$. The region $\mathcal{C}_i$ is best in the sense that subject to $\alpha_i \leq \alpha_i'$, where α_i' is some maximum tolerable error for H_i and we may have $\alpha_i' = \alpha$ for all i, $\mathcal{C}_i$ maximizes β_i. The set of symptoms to be observed must be determined so that $\beta_i \geq \beta_i'$, where β_i' is some minimum power for the test of H_i and, again, we may have $\beta_i' = \beta$ for all i. In addition to satisfying this minimum power requirement, the set of symptoms to be observed must satisfy the restrictions discussed with respect to maximizing the number of classes.

Estimates of all probabilities, $Pr(x_i|C_i)$ and $Pr(x_i|C_i^{-1})$ for all possible observations x_i and for $i = 1,2,\ldots,m$, are required and must be obtained for a large sample of the population. Since we require probabilities of the various configurations in the populations of patients belonging to each of the m classes, C_i, and in the populations which are the complements, C_i^{-1}, of these m classes, it is imperative that there be other methods of diagnosis, whose verdict is taken to be correct, but which may be considered infeasible for inclusion in the multiphasic screening examination. In such instances, necessary follow-up to provide supplemental diagnostic examinations must be made on all persons in the sample to give a final diagnosis and classification into the various disease classes. Because of the time lag necessary to achieve this final diagnosis and the low prevalence of many diseases to be considered, it seems reasonable that observations may have to be taken not only from the proposed multiphasic examination population but also from the populations consisting of previously diagnosed cases. Studies will also be required on the homogeniety of the population to which the examination is to be given to determine whether such factors as age will require initial classification of the population before classification according to final diagnosis in obtaining estimates of the probabilities.

There is also assumed a standardization and quantification of the various symptoms. For example, more information is required from the patient questionnaire than "Have you had a recent unexplained weight loss?" How much weight lost over what period of time might be a reasonable inquiry where certain combinations of time and amount might be considered significant. Furthermore, even on symptoms which are necessarily quantitative, a value must be determined to delineate normal and abnormal and hence to indicate when a symptom measurement is negative $(X_{ij} = 0)$ and when it is positive $(X_{ij} = 1)$.

Having established the m hypotheses $H_1 : C_1, \ldots, H_m : C_m$ with corresponding alternative $K_1 : C_1^{-1}, \ldots, K_m : C_m^{-1}$, critical rejection regions $\mathcal{C}_1, \ldots, \mathcal{C}_m$, significance levels $\alpha_1, \ldots, \alpha_m$, and powers $\beta_1, \ldots, \beta_m$, consider simultaneous testing of these m hypotheses. There are 2^m theoretically possible true situations and 2^m sequences of decisions resulting from testing the m hypotheses. Denote by $\phi_i = (\phi_{i1}, \phi_{i2}, \ldots, \phi_{im})$, $i = 1, \ldots, 2^m$, the possible m-vectors consisting of 1's and -1's, and by $H_i' : \prod_{j=1}^{m} C_j^{\phi_{ij}}$ the composite hypothesis that those hypotheses H_j for which $\phi_{ij} = 1$ are true and those for which $\phi_{ij} = -1$ are false. Then H_i' is to be accepted if and only if each H_j with $\phi_{ij} = 1$ is accepted and each H_j with $\phi_{ij} = -1$ is rejected. For example, let $m = 2$, $\phi_i = (1, -1)$ so that $\phi_{i1} = 1$ and $\phi_{i2} = -1$; then $H_i' : C_1 C_2^{-1}$ denotes the hypothesis that the patient belongs to class C_1 but not to class C_2, and we would accept H_i' if and only if we accepted $H_1 : C_1$ and rejected $H_2 : C_2$. Ideally, one would want estimates of $\Pr(X_{11}, \ldots, X_{1n_1}, \ldots, X_{m1}, \ldots, X_{mn_m} \mid \prod_{j=1}^{m} C_j^{\phi_{ij}})$ for each of the 2^m disease states and critical regions of rejection for each of the $2^m(2^m - 1)/2$ cases $H_i' : \prod_{j=1}^{m} C_j^{\phi_{ij}}$ vs. $H_k' : \prod_{j=1}^{m} C_j^{\phi_{kj}}$. In view of the magnitude of the minimal number, m, of classes, the above approximation is suggested, even though in practical application, prevalence considerations would allow reduction of the number of disease class combinations to be considered.

The probabilities of error on each hypothesis H_j will eventually be known, but there is still the question of the total probability of error when the m hypotheses are compounded. Suppose that the m two-decision problems are independent. The probability of accepting H_i' when H_k' is actually true becomes

$$\Pr(H_i'|H_k') = \prod_{j=1}^{m} (1-\alpha_j)^{(1+\phi_{ij})(1+\phi_{kj})/4} (1-\beta_j)^{(1+\phi_{ij})(1-\phi_{kj})/4} \cdot$$

$$\alpha_j^{(1-\phi_{ij})(1+\phi_{kj})/4} \beta_j^{(1-\phi_{ij})(1-\phi_{kj})/4} .$$

In particular, the probability of a correct decision is

$$\Pr(H_i'|H_i') = \prod_{j=1}^{m} (1-\alpha_j)^{(1+\phi_{ij})/2} \beta_j^{(1-\phi_{ij})/2} .$$

Defining a healthy person as one for whom all m hypotheses are false,

$$\phi_{ij} = -1 , \quad j = 1,\ldots,m$$

$$\Pr\left(\prod_{j=1}^{m} C_j^{-1} \,\middle|\, \prod_{j=1}^{m} C_j^{-1}\right) = \prod_{j=1}^{m} \beta_j ,$$

so that the probability of doing further study in at least one disease class C_i for a person who is healthy is

$$1 - \Pr\left(\prod_{j=1}^{m} C_j^{-1} \,\middle|\, \prod_{j=1}^{m} C_j^{-1}\right) = 1 - \prod_{j=1}^{m} \beta_j .$$

This last probability is a special case of the probability of at least one wrong decision when H_i' is true, $1 - \Pr(H_i'|H_i') = 1 - \prod_{j=1}^{m} (1-\alpha_j)^{(1+\phi_{ij})/2} \cdot \beta_j^{(1-\phi_{ij})/2}$.

Further study is required to determine conditions under which the m tests will be independent as well as to determine whether such conditions will be satisfied in this application. In the case it is not reasonable to assume independent decisions on each of the hypotheses H_j, then the whole question of total error when compounding probabilities remains unanswered. Further study and discussion are necessary also 1) to provide a systematic and statistically valid method for selecting the symptoms to be observed for each disease class 2) to determine the additional error induced in the decisions by the fact that we know only random estimates of the desired parameters, 3) to estimate the total probability of a wrong decision on each test H_j vs. K_j ; i.e., $Pr(X_i \in \mathcal{C}_i | C_i)\ Pr(C_i) + Pr(X_i \notin \mathcal{C}_i | C_i^{-1})[1 - Pr(C_i)] = \alpha_i p_i + (1 - \beta_i)(1 - p_i)$ where p_i is the prevalence of disease class C_i.

The reader is referred to the following papers which have been helpful in reaching the formulation of the problem as presented in this report. The theory of testing a simple hypothesis against a simple alternative is treated in the textbook by Neyman [9], and the examples include a discussion of screening for tuberculosis. The studies of Chiang [5], Neyman [8], Taylor [10], and Yerushalmy [12], point out interesting problems in obtaining estimates of probabilities in the field of public health as well as illustrate the theory of estimation due to Neyman [7]. The excellent paper by Chiang, Hodges, and Yerushalmy [4], discusses in a very general way several applications of statistics to medical diagnosis. The two papers on the problem of classification — Anderson [1], and Birnbaum [3] — are pertinent to this study since they both involve deciding from which population a person comes on the basis of his vector $(X_1, \ldots, X_p)$ of observed symptoms or traits. The two-decision problems discussed in this report are but a special case of the k-decision problems considered by Anderson or Birnbaum. Finally, the problem of generating a complex statistical test by simultaneous consideration of more simple testing problems is given theoretical treatment by Lehmann in [6]. While his paper does not discuss the question of total error, the works of Birnbaum [2] and Wallis [11] do discuss error in compounding tests, although in a different context from the application in this study.

BIBLIOGRAPHY AND SUGGESTED REFERENCES

1. Anderson, T.W., "Classification by Multivariate Analysis," Psychometrika, Vol. 16, No. 1, (March 1951).

2. Birnbaum, Allan, "Combining Independent Tests of Significance," Journal of the American Statistical Association, Vol. 49, (Sept. 1954) pp. 559-574.

3. Birnbaum, Allan, and A.E. Maxwell, "Classification Procedures Based on Bayes's Formula," Applied Statistics, Vol. 9, No. 3, (Nov. 1960).

4. Chiang, Chin Long, J.L. Hodges, and Jacob Yerushalmy, "Statistical Problems in Medical Diagnoses," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, (1955).

5. Chiang, Chin Long, "On the Design of Mass Medical Surveys," Human Biology, Vol. 23, No. 3, (Sept. 1951).

6. Lehmann, E.L., "A Theory of Some Multiple Decision Problems I. Simultaneous Testing of Several Hypotheses," Annals of Mathematical Statistics, Vol. 28, No. 1, (1957).

7. Neyman, Jerzy, "Contribution to the Theory of the χ^2 Test," Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability, (1946).

8. Neyman, Jerzy, "Outline of Statistical Treatment of Diagnosis," Public Health Reports, Vol. 62, No. 40 (Oct. 1947).

9. Neyman, Jerzy, First Course in Probability and Statistics, Henry Holt, New York, 1950.

10. Taylor, William F., "Mathematical Statistical Studies on the Tuberculin and Histoplasmin Skin Tests," Human Biology, Vol. 23, No. 1 (Feb. 1951).

11. Wallis, W. Allen, "Compounding Probabilities from Independent Significance Tests," Econometrica, Vol. 10, Nos. 3-4 (1942).

12. Yerushalmy, Jacob, "Statistical Problems in Assessing Methods of Medical Diagnosis," Public Health Reports, Vol. 62, No. 40, (Oct. 1947).

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

Head, Logistics and Mathematical
Statistics Branch
Office of Naval Research
Washington 25, D.C.

C.O., ONR Branch Office
Navy No. 100, Box 39, F.P.O.
New York City, New York

ASTIA Document Service Center
Arlington Hall Station
Arlington 12, Virginia

Institute for Defense Analyses
Communications Research Div.
von Neumann Hall
Princeton, New Jersey

Technical Information Officer
Naval Research Laboratory
Washington 25, D.C.

C.O., ONR Branch Office
346 Broadway, New York 13, NY
Attn: J. Laderman

C.O., ONR Branch Office
1030 East Green Street
Pasadena 1, California
Attn: Dr. A.R. Laufer

Bureau of Supplies and Accounts
Code OW, Dept. of the Navy
Washington 25, D.C.

Professor Russell Ackoff
Operations Research Group
Case Institute of Technology
Cleveland 6, Ohio

Professor Kenneth J. Arrow
Serra House, Stanford University
Stanford, California

Professor G.L. Bach
Carnegie Institute of Technology
Planning and Control of Industrial
Operations, Schenley Park
Pittsburgh 13, Pennsylvania

Professor A. Charnes
The Technological Institute
Northwestern University
Evanston, Illinois

Professor L. W. Cohen
Math. Dept., University of Maryland
College Park, Maryland

Professor Donald Eckman
Director, Systems Research Center
Case Institute of Technology
Cleveland, Ohio

Professor Lawrence E. Fouraker
Department of Economics
The Pennsylvania State University
State College, Pennsylvania

Professor David Gale
Dept. of Math., Brown University
Providence 12, Rhode Island

Dr. Murray Geisler
The RAND Corporation
1700 Main Street
Santa Monica, California

Professor L. Hurwicz
School of Business Administration
University of Minnesota
Minneapolis 14, Minnesota

Professor James R. Jackson
Management Sciences Research
Project, Univ. of California
Los Angeles 24, California

Professor Samuel Karlin
Math. Dept., Stanford University
Stanford, California

Professor C.E. Lemke
Dept. of Mathematics
Rensselaer Polytechnic Institute
Troy, New York

Professor W.H. Marlow
Logistics Research Project
The George Washington University
707 - 22nd Street, N.W.
Washington 7, D.C.

Professor Oskar Morgenstern
Economics Research Project
Princeton University
92 A Nassau Street
Princeton, New Jersey

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

Professor R. Radner
Department of Economics
University of California
Berkeley, California

Professor Stanley Reiter
Department of Economics
Purdue University
Lafayette, Indiana

Professor Murray Rosenblatt
Department of Mathematics
Brown University
Providence 12, Rhode Island

Mr. J. R. Simpson
Bureau of Supplies and Accounts
Navy Department (Code W31)
Washington 25, D.C.

Professor A. W. Tucker
Department of Mathematics
Princeton University
Princeton, New Jersey

Professor J. Wolfowitz
Department of Mathematics
Lincoln Hall, Cornell University
Ithaca 1, New York