

SPECIAL ARTICLE

REPORT ON MATHEMATICS IN THE MEDICAL SCIENCES*

DAVID D. RUTSTEIN, M.D.,† MURRAY EDEN, PH.D.,‡ AND MARCEL P. SCHÜTZENBERGER, M.D.§

BOSTON

THIS conference on "Mathematics in the Medical Sciences" was called because medical scientists have begun to ask questions that entail the development of new mathematical theory and the application of more complex mathematical reasoning than has been the case in the past. Also, the effective use of new instruments now becoming available to medicine demands an understanding of their underlying physical theory and its mathematical application. The meeting was concerned with both the nature of mathematics and the technics for its application that have been useful to biology and medicine in the past or may be applicable in the near future. In this report — which is not a complete summary of the meeting — the lectures and discussions are focused on the principles governing the interrelation of mathematical theory and the biologic and medical sciences. After the summary there is a general statement by Professor William G. Cochran that epitomizes the present status of "The Role of Mathematics in the Medical Sciences."

For the purposes of this report, biomathematics is defined as the development of new mathematical theories or technics under the stimulus of unsolved biologic problems and the application of existing mathematical theory and technics for describing and interpreting biologic and medical phenomena.

There is considerable overlap between the fields of biomathematics and statistics. To identify biomathematics more clearly, the field of statistics is briefly discussed, and relations to biomathematics are indicated. Biomathematics as considered at the Conference will then be presented, with particular reference to the construction of hypotheses or "model

building."

STATISTICS

Much of the statistical work in the medical sciences in the recent past has been concerned with the systematic application of logical reasoning to the solution of particular quantitative problems. Well known are such applications as drug assays, testing of laboratory procedures and sample and mass surveys. As this area of statistics has developed in relation to medical research, there has been increasing interest in the principles governing the design and analysis of experiments. An appropriate design appears to be crucially important, especially in problems involving large numbers of investigators. Moreover, precise analysis and interpretation of the data are the last important steps of a successful experiment. This kind of statistics has become classic and was intentionally omitted from the program.

But before this use of statistics is dismissed, one relation to biomathematics is worth mentioning. The use of human subjects imposes limitations on the design of experiments or makes human experimentation impossible. Sometimes questions can be restated so that controlled experiments involving human beings can be done. For instance, it may be impossible to perform such an experiment to determine whether a new drug is better than no treatment in a disease like rheumatic fever, but it is possible to perform a controlled experiment on whether the new drug is better than the generally accepted treatment. But when human experiments dependent upon classic statistics cannot be performed, as in experiments involving human genetics, mathematical elaborations of Mendelian theory may make it possible to predict human genetic distributions.

Another relation between statistics and biomathematics is concerned with the application of statistics in the exploration of a new field of investigation or a new set of problems. In this sense, statistics has a use analogous to that of the microscope or dye permitting the visualization of details not perceptible to the naked eye. Three speakers at the Conference presented illustrations of this use of statistics. Pipberger described a method of principal component analysis of electrocardiographic tracings and a procedure for discriminating between tracings according to some particular property. Thus, with the help of time integrals and spatial rotations, it was possible to identify the elements of the electrocardiogram that correlated independently with such abnormalities as ventricular hypertrophies and myocardial

*Based on a conference held under the auspices of the Department of Preventive Medicine, Harvard Medical School, Boston, January 16 and 17, 1961. Program titles and participants were as follows: "Order of Amino Acids in a Protein," Sidney A. Bernhard, National Institute of Mental Health; "Genetic Mapping," Cyrus Levinthal, Massachusetts Institute of Technology; "Mathematical Models for Muscular Contraction," Richard Podolsky, United States Naval Medical Research Institute; "Mathematical Models for the Study of Physiological Systems," Norman Z. Shapiro, National Institutes of Health; "The Role of Mathematics in the Medical Sciences," William G. Cochran, Harvard University; "Mathematical Models and Computational Technics for the Analysis of Neuroelectric Activity," Walter A. Rosenblith, Massachusetts Institute of Technology; "Collection, Storage, Analysis and Use of Electrocardiographic Data," H. V. Pipberger, Georgetown University School of Medicine and United States Veterans Administration; "Potential Applications of Mathematics to the Diagnosis of Illness," Murray Eden, Massachusetts Institute of Technology; and "Use of Mathematics in the Study of Biologic Transport Systems," John L. Stephenson, National Heart Institute.

We are grateful to Professor Norman F. Ramsey, of Harvard University, Professor John R. Pappenheimer, of Harvard Medical School, and Dr. James A. Shannon, of the National Institutes of Health, for their contributions to the Conference as chairmen of the sessions.

†Professor of preventive medicine and head, Department of Preventive Medicine, Harvard Medical School.

‡Associate professor of electrical engineering, Center for Communications Sciences, Research Laboratory of Electronics, Massachusetts Institute of Technology; lecturer on preventive medicine, Harvard Medical School.

§Maître de conférences, Faculty of Sciences, University of Poitiers, Poitiers, France; visiting professor, Department of Statistics, University of North Carolina.

infarction, and eliminate those that were dependent and introduced no new evidence of such involvement.

Rosenblith, in a problem of pattern recognition, described the use of the statistical theory of communication as a means of exhibiting regularities in electroencephalographic tracings and in recordings of electrical events observed in the nervous system.

Eden presented a third application of this use of statistics. This concerned the striking of a proper balance between information retrieval and the automation of various aspects of medical diagnosis and treatment. Thus, methods exist for the retrieval of such items of information as laboratory reports and other quantitative data, whereas the defining, storing and retrieving of descriptive data such as that found in the history and physical examination are as yet unsolved. Eden emphasized the fact that, in the final analysis, machine diagnosis of illness will depend on information concerning the probability of specific symptoms, signs or laboratory tests appearing in certain diseases, and that to be applicable, these data will have to be collected in many kinds of population groups.

These newer applications of statistics would be prohibitively time consuming were it not for recent technologic development in the computation art. First of all, very large amounts of data can be processed rapidly. Secondly, it may be possible to build a simple, special-purpose computer that will perform the preliminary processing of the data for the clinician. If further analysis is desirable, this first step will have been helpful in preparing the data for introduction into a large, general-purpose computer.

It was made clear throughout the Conference that computers can be helpful only when the investigator has carefully defined his terms, meticulously planned the design of his experiments, established a logical method of classification and collected his information in consistent fashion.

THE CONSTRUCTION OF HYPOTHESES OR "MODEL BUILDING"

Much of the time of the Conference was devoted to the use of mathematics in the construction of hypotheses useful in medical and biologic research. This involves the analysis of existing data with the help of mathematical theory so that more useful hypotheses may be created than would otherwise be the case. This use of mathematics in constructing biologic or medical hypotheses is often referred to as model building. Indeed, the creation of mathematical models has been implicit in all the statistical applications mentioned above because the particular probability theory upon which the statistics are based in itself represents a mathematical model.

Mathematical models are particularly useful when the investigator by intuitive means cannot see the relation between two or more facts or sets of data. In such a situation the mathematician may establish a deductive relation through a mathematical equa-

tion that can then be tested by experiment. An example in a physical system is Kundt's deduction that the velocity of sound in a gas is related to its specific heats. This was later verified by experiment. An example in biology proposed by Shapiro concerns the use of labeled compounds in the study of metabolism. If one uses a mathematical model involving simultaneous linear differential equations with constant coefficients for a description of the kinetics of a labeled system, one can mathematically deduce the form of relation between amounts of labeled compounds measured and time — a form that is subject to empirical verification.

PHYSICAL MODELS

Mathematicians may also be helpful by pointing out that certain well studied physical systems may be useful in the investigation of particular biologic systems. The extensive use of physical models is supported by the ideas that physicists have been successful in explaining nature, that alternatives to the vitalist approach must be physical ones and that the mathematical knowledge of the average biologist is limited to that used in simple physical systems. Where a physical system is applicable, this approach may be successful. Thus, the Hodgkin-Huxley model of nerve transmission is based on electromagnetic theory and that of the transmission of electric power through cables.

There are, however, limitations to the use of physical systems in biologic research. Biologic measurements often cannot be made with the great precision of physical ones. Moreover, the clearly defined conditions of equilibrium in physical systems often cannot be approximated in biology, even when static measurements are precise. Thus, the Russians could calculate the movements of the Earth and Venus, the trajectory of the carrier rocket and the point, direction and time of firing of the interplanetary rocket on its way to Venus. In biologic research, the chloride concentration of the blood may be measured with some precision at a particular time, but because of the many known and unknown factors that may influence it, it may not be possible to predict its value from moment to moment, or in different compartments of the body.

Furthermore, physical systems can often be simplified without disturbing the realistic relation of the simplified to the original system. In contrast, it may be very difficult in biology to simplify a system without making it totally artificial. Thus, the Donnan equilibrium may be measured precisely in a simplified physical system, but its direct application to the formation of cerebrospinal fluid might be unrealistic.

Where it is impossible to find a simple physical model that is directly applicable to a biologic system, mathematical theory may be helpful. Thus, Stephenson showed that all the kinds of data needed on the basis of a physical model to study the complex prob-

lem of fat absorption cannot be collected in a human subject. But it was possible with fatty acid tracer data and the use of integral equations to work out the absorption, splitting and recirculation of fatty acids and of triglycerides. This approach may be very useful in the study of metabolism and in other biologic and medical problems. In this connection, it is instructive that the type of equation used by Stephenson is known to mathematicians as the Volterra equation and has a broad range of application to biologic and physical systems. Actually, Volterra was a mathematician who developed this equation for the study of another biologic problem — that concerned with the competition of species.

MONTE CARLO METHODS

Until recently, mathematical models of complex biologic systems could not be verified because the mathematics required to get to the experimental implementations of the models is remarkably complicated and the computations inordinately long. Such a model was presented by Podolsky with reference to the kinetics of muscular contraction. He could study his model because the large, general-purpose digital computer can by reason of its great speed do what a human computer would find prohibitively time consuming. But even this approach is limited. Some biomathematical models, although made up of simple elements, become complex because they have so many interactions that even the largest computer cannot by analytic methods produce a specific answer for the entire model. Thus, when Podolsky's model simulated the transition from one steady state of muscular contraction to another, a new approach was necessary. Here, Podolsky and Shapiro used Monte Carlo methods, by which, in terms of a statement of probability, they ran a series of mathematical experiments on a digital computer and were able to converge on a numerical solution to the entire problem. Monte Carlo methods are presently under study by a number of mathematicians, and are applicable to many biologic and medical problems.

STOCHASTIC PROCESSES

In many biologic systems there is no present possibility of making a physical model, however complicated. Instead, certain factors must be regarded as having values controlled by chance. Processes of this kind are referred to as stochastic. Stochastic mathematical models are models in which the behavior of the system under study is modified by the addition of a variable with a known probability of occurrence and distribution. *At the Conference*, Rosenblith demonstrated the use of such a model by the effect of a repetitive standardized auditory stimulus on the neuroelectric activity of the human brain. Such models have also been used in many kinds of

biologic studies, including epidemiology by Bailey, in cancer induction by Neyman and evolutionary genetics assessing a long-range possibility of eugenics by Dahlberg. This approach, supported by computer facilities, will enormously broaden the opportunities of the application of mathematics to the study of biologic and medical phenomena.

FINITE MATHEMATICS

In applying mathematics to biology, biologists must not limit their conception of mathematics to a kind of glorified arithmetic concerned only with the relation between quantities of things. Actually, as illustrated at the Conference, a good deal of modern mathematics is concerned with the relations between objects rather than with numerical values. Bernhard applied this kind of mathematics to the sequencing of amino acids in a protein. In collaboration with Duda, he developed a logical procedure for determining the sequence by digital computation. This approach has come to be called finite mathematics and has been used in other biologic research — for example, by McCulloch and Pitts in their hypothesis of the random nerve net to explain the mechanism of the nervous system. This procedure can also be used for other than biologic problems, as was demonstrated by Eden at the Conference in his synthesis of human handwriting by mathematical methods. The stress in finite mathematics is on combinatorial technics. Levinthal discussed the combinatorial procedures involved in genetic mapping and posed certain unanswered questions relating to the verification of the assumption of the linearity of the genetic map. This same mathematical approach has been used by the team of Levi-Strauss, a biologist, and Weil, a mathematician, in studying laws governing marriage relations in primitive societies, and illustrates the application of this method to the study of anthropology.

SPECIFICITY OF HYPOTHESES

Shapiro made a significant point when he declared that a mathematical model (hypothesis) to be useful had to be stated in specific and restrictive terms. If the hypothesis when tested by experiment can be satisfied by many different sets of data, it is too general and has little differential or predictive value. As an example Shapiro showed that the multilinear receptor theory of color vision* was satisfied by many sets of data collected when observers matched lights in the entire range of the visible spectrum.

INFORMATION THEORY AND CYBERNETICS

Except for a brief reference on neuroelectric phenomena by Rosenblith, the application to biology of the mathematics concerned with information

*This theory states that color vision in the human eye depends upon a finite number (three?) of receptors, each of which follows Beer's law.

theory and cybernetics was not covered in the Conference but should be mentioned briefly here. At a time when biologists, attempting to bring rigor into their field of interest, felt obliged to adhere to quantitative physical models, it remained for the mathematician Norbert Wiener to point out that the way the organism processes energy may be much less important than the mechanisms by which these processes are controlled. The concept of information theory and the feedback of cybernetics have also been applied to the study of perception by Reichardt in the vision of the bee, and by Stark in studies of muscular and pupillary contractions.

APPLICATION OF STANDARD MODELS

Because of limitations of time, the program and the meeting were devoted almost entirely to questions requiring both biologic and mathematical research. Only brief mention was made of the application of standard mathematical methods to problems in biology, in medicine and in medical care. If more time had been available, presently unexploited applications of standard mathematical methods would have been discussed. For example, scientists in the Department of Statistics of the University of North Carolina have recently applied mathematical techniques used in the "inventory control problem" to the operation of a blood bank — that is, the minimum number of units of blood in a bank required to meet all essential needs.

EPILOGUE

Some of the ways in which the life sciences and mathematics can interact were outlined at the Conference. The notions that mathematics can only be used as a service tool in biology or that biology was either too vague or too trivial to engage the interest of the mathematician were dispelled. It also became obvious that most medical and biologic investigators know too little mathematics of either the classical or modern variety to work productively with the mathematician. On the other hand, it became evident that mathematical theory and technics need extensive development for the effective study of biologic problems. Ideally, prospective life scientists should be well grounded in both biology and in mathematics, but such persons will always be rare. In spite of this, fruitful collaboration can go on between the two disciplines if biologic and medical investigators will learn enough mathematics, electrical engineering and physics to be able to ask the proper questions of the mathematician, and if the mathematician will learn enough biology to be able to develop the mathematical theory and technics necessary to the solution of important biologic and medical problems.

SUGGESTED BIBLIOGRAPHY

General

Oncley, J. L. *Biophysical Science*. 609 pp. New York: Wiley, 1959.

- Wiener, N. *Cybernetics*. 194 pp. New York: Wiley, 1948.
- Contributions to Biology and Problems of Health: Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability, v.4: Held at the Statistical Laboratory, University of California, December, 1954, July and August, 1955*. Edited by J. Neyman. 187 pp. Berkeley: Univ. of California Press, 1956.
- Contributions to Biology and Problems of Health: Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, v.5: Held at the Statistical Laboratory, University of California, Sept. 12th-16th, 1955*. Edited by J. Neyman. Berkeley: Univ. of California Press (in press).
- Rashevsky, N. *Mathematical Biophysics: Physico-mathematical foundations of biology*. Third edition. 2 vol. New York: Dover, 1959-1960.
- Symposium on Information Theory, London, 1955. *Information Theory: Papers read at Symposium held at the Royal Institute, London, Sept. 12th-16th, 1955*. Edited by C. Cherry. 401 pp. London: Academic Press, 1956.
- Symposium on Information Theory, London. *Information Theory: Papers read at Symposium held at the Royal Institute, London, Aug. 30-Sept. 4, 1960*. Edited by C. Cherry (in press).
- Interdisciplinary Conference on Self-Organizing Systems, Chicago, 1959. *Self-Organizing Systems: Proceedings*. Edited by M. C. Yovits and S. Cameron. 322 pp. New York: Pergamon, 1960. (*International Tracts in Computer Science and Technology and Their Application*. Vol. 2.)
- Symposium on Information Theory in Biology: Held at Gatlinburg, Tennessee, October 29, 1956*. Edited by H. P. Yockey, H. Quastler and R. Platzman. 418 pp. New York: Pergamon, 1958.
- Bailey, N. T. J. *Mathematical Theory of Epidemics*. 194 pp. New York: Hafner, 1957.

Genetics and Protein Structure

- Benz, S. On topography of genetic fine structure. *Proc. Nat. Acad. Sc.* **47**:403, 1961.
- Levinthal, C. Coding aspects of protein synthesis. *Rev. Mod. Physics* **31** (1):249-255, 1959.
- Levinthal, C. Genetic and chemical studies with alkaline phosphatase of *E. coli*. *Brookhaven Symposia in Biol.* **12**:76-85, 1959.
- Levinthal, C., Garen, A., and Rothman, F. Relationship of gene structure to protein structure: studies on alkaline phosphatase of *E. coli*. In *Fifth International Congress of Biochemistry: Moscow, 1961* (in press).
- Bernhard, S. A. Simple model of molecular specificity in enzyme-substrate systems. I. Theory and applications to system acetylcholinesterase-substrate. *J. Am. Chem. Soc.* **77**:1966-1972, 1955.
- Bernhard, S. A. Simple model of molecular specificity in enzyme-substrate systems. II. Correlation of Michaelis constant with inhibition constant. *J. Am. Chem. Soc.* **77**:1973, 1955.
- Bernhard, S. A. Some mathematical consequences of interaction of enzyme with two reactant species. *Disc. Faraday Soc.* **20**:267, 1956.
- Bernhard, S. A. Symposium on protein microstructure. *J. Polymer Sc.* (in press).
- Bernhard, S. A., Katchalski, E., Berger, A., and Sela, M. Polyfunctional catalysis in polypeptides and enzymes. *J. Am. Chem. Soc.* (in press).

Mathematical Models (Physiology)

- Podolsky, R. J. Nature of contractile mechanism of muscle. In *Biophysics of Physiological and Pharmacological Techniques*. Edited by A. M. Shanes. Washington, D. C.: American Association for the Advancement of Science (in press).
- Shapiro, N., and Vreenegeer, H. Generalized important event technique. *J. A. Comp. Machinery* (in press).
- Stephenson, J. L. Theory of measurement of blood flow by dilution of indicator. *Bull. math. Biophys.* **10**:117-121, 1948.
- Stephenson, J. L. Theory of measurement of blood flow by dye dilution technique. *IRE Tr.*, pp. 81-88, December, 1958. (PGME-12.)
- Stephenson, J. L. Theory of transport in linear biological systems. I. Fundamental integration equation. *Bull. math. Biophys.* **22**:1-17, 1960.
- Stephenson, J. L. Theory of transport in linear biological systems. II. Multiflux problems. *Bull. math. Biophys.* **22**:113-138, 1960.

Medical Diagnosis

- Proceedings of conference on diagnostic data processing. *IRE Tr.*, pp. 232-368, October, 1960. (ME-7.)

Neurophysiology

- Communications Biophysics Group of Research Laboratory of Electronics, and Siebert, W. M. *Processing Neuroelectric Data*. Cambridge, Massachusetts: Massachusetts Institute of Technology, July 7, 1959. (Tech. Rep. No. 351.)
- Barlow, J. S. Rhythmic activity induced by photic stimulation in relation to intrinsic alpha activity of brain in man. *Electroencephalog. & Clin. Neurophysiol.* **12**:317-326, 1960.
- Clark, W. A., et al. Average response computer (ARC): digital device for computing averages and amplitude and time histograms of electrophysiological response. *IRE Tr.*, pp. 46-51, January, 1961. (BME-8.)
- Gerstein, G. L., and Kiang, N. Y.-S. Approach to quantitative analysis of electrophysiological data from single neurons. *Biophys. J.* **1**:15-28, 1960.
- Goldstein, M. H. Statistical model for interpreting neuro-electric responses. *Information & Control* **3**:1-17, 1960.
- Peake, W. T. *An Analytical Study of Electric Responses at the Periphery of the Auditory System*. Cambridge, Massachusetts, Research Laboratory of Electronics, March 17, 1960. (RLE Tech. Rep. No. 365.)
- Farley, B. G., Frishkopf, L. S., Clark, W. A., Jr., and Gilmore, J. T., Jr. *Computer Techniques for the Study of Patterns in the Electroencephalogram*. Technology Press, Cambridge, Massachusetts, Research Laboratory of Electronics, November 6, 1957. (RLE Tech. Rep. No. 337, & Lincoln Lab. Tech. Rep. No. 165.)