The study presented here gives examples of the use of mathematical models in the investigation of the epidemiological trend of tuberculosis. These models are based on methods developed and utilized in other sciences.

THE USE OF MATHEMATICAL MODELS IN THE STUDY OF THE EPIDEMIOLOGY OF TUBERCULOSIS

Hans Waaler, cand. oecon.; Anton Geser, M.D.; and Stig Andersen, Cand. Polit.

A.J.P.H 52, Nº 6: 1002, 1962.

IS THE epidemic wave of tuberculosis in India and other underprivileged countries on its way up or down? What is the possible impact on the tuberculosis problem of control programs in these countries?

It is the contention of this paper that such vital epidemiological questions may not have to remain unanswered for years. The future trend is determined by the dynamics inherent in the tuberculosis situation of today and can be extracted if the appropriate technics are developed.

The epidemiological factors, such as prevalence and incidence of infection and disease, are mutually dependent and between them determine the trend of tuberculosis. These factors are closely linked together in a set of relationships. The potentiality of available data has not always been fully utilized in the field of epidemiology; extensive data collection has been and is going on with little attempt to project the data into the future.

Other sciences such as meteorology, demography, and in particular, economics, have realized more fully that it is possible by an intelligent combination of all the relevant factors which are operating on the problem under study, to predict something about the future trend. This realization has led economists and other scientists to construct mathematical models in which the quantified relevant factors are combined in mathematical relationships, and such models have substantially contributed to the deeper understanding of the problems.

A priori it would appear more promising to utilize the model approach in the field of tuberculosis epidemiology than in the field of economics and other social sciences, where the operating agent is the unpredictable homo sapiens. The economist must realize that his popualtion may, for instance, deliberately change its buying and selling habits and the demographer must be prepared to face deliberate attempts by his population to change the birth rate.

Notwithstanding the problems connected with the development of resistance to certain chemotherapeutics, the epidemiologist, on the other hand, hardly has to fear that the tubercle bacilli will deliberately change their virulence or infectivity even if they were losing ground, failing to propagate themselves rapidly enough to maintain their population size.

In the field of epidemiology the model approach has been attempted in the study of acute diseases by several workers. Hedrich,[1] for instance, was able to predict waves of measles by calculating the number of susceptibles at various points of time. Recently

Muench[2] has developed to a high degree of perfection certain aspects of uses of models in epidemiology. He studies the validity of certain hypotheses regarding mechanisms of infection by comparing results of his models to actual observations.

In the epidemiology of tuberculosis, the model approach was first applied by Frost.[3] He predicted that with the low and falling rate of transmission of tuberculous infection in the United States, it was most likely that the tubercle bacilli were fighting a losing battle, since they would be unable to transmit sufficient infection to maintain the balance in their own favor.

In 1957, Feldman,[4] by further progress along the road of model-making, was able to confirm the predictions of Frost.

The model approach is also implied in a work by Palmer, Shaw, and Comstock[5] where they make predictions regarding the impact of a BCG campaign under certain assumed epidemiological conditions. Sutherland[6] also uses model thinking in a publication on the British Medical Research Council's BCG trial where he tries to estimate the reduction in the incidence and prevalence of tuberculosis, which could have been achieved in England and Wales if BCG vaccine had been given to all school-leavers at the time of the trial.

The model approach has more recently been applied by Frimodt-Møller[7] in connection with a longitudinal tuberculosis survey in South India. In this survey Frimodt-Møller obtained estimates of most of the necessary epidemiological data through a long-term follow-up in a large population group. He then applied these data to a hypothetical population, introduced the effect of treatment, and calculated the outcome of the situation in order to evaluate the effect of the control program.

The present authors believe it would be possible and fruitful to develop the model approach further by more explicitly applying mathematical statistical methods from other sciences. This would necessitate the closest cooperation between epidemiologist and statistician, since an intimate knowledge of the epidemiology of tuberculosis as well as of the mathematical methods are essential. This paper is a result of such cooperation and purports to show how a simple epidemetric model, reflecting the dynamics of tuberculosis, can be constructed. Examples showing the potentialities of the model in giving the time trend of tuberculosis and in evaluating control programs are also given.

Definition and Terminology

The term Model in its broader sense merely means a systematic way of organizing and quantifying factors relevant to the analysis of a complex problem. The construction of a model involves the formulation of a hypothesis about the nature of the relationships that exist between the various relevant factors. The validity of the hypothesis can be tested by feeding data into the model, solving it through calculations, and finally comparing the results with actual observations. This requires, of course, that the underlying assumptions on which the model is constructed do not significantly change during the period under consideration.

If, for instance, the communication intensity is included as a variable, then the relationships are valid irrespective of time trends in communication intensity. If not included as a variable, the validity rests on the assumption of constant intensity of communication. This is what is usually referred to as the problem of autonomy of the relationships.

One of the merits of the model approach is that it helps to orient the

thoughts in a systematic way and at every step clearly defines the assumptions made. In fact, a model is a quantified simplification of the bewildering complexity of reality. Model-thinking is in many ways analogous to map-making. To be useful, the map must disregard many of the details of the landscape and only retain those features which will serve the purpose for which the map is made, e.g., to guide the construction of a road or a journey through a continent. So with the model: only factors relevant to the purpose for which the model is made should be considered. It is the task of the epidemiologist to define the relevant mechanisms to be included in his model.

In a more restricted sense, a model is defined as a set of mathematical relationships which together give the underlying mechanisms and which, when solved, will give a quantified description of the problem. Static models are constructed to give the picture of the problem at a given point of time. In this paper we will, however, only deal with the generally more useful dynamic models. A dynamic model can briefly be defined as a model in which more than one point of time enters the calculations. These models will thus show how the variables change over a period of time, or in other words, the trend of the variables.

Certain variables which are the quantities relevant to the problem enter into the construction of the model. In TB epidemiology, prevalence and incidence of infection and disease and TB mortality are obviously relevant variables. These can be conveniently subdivided into two groups characterized by their statistical properties, viz., stock quantities (existing at a given point of time) and flow quantities (giving the flow out and in of the stock quantities). In the field of epidemiology stock quantities are represented by prevalences, and flow quantities by incidences.

In addition to the variables, another category of quantities enters the models. These are the parameters or constants that define the form and content of the relationships.*

The solution of the model is achieved by feeding it with a set of observed or estimated epidemiological data pertaining to a given area at a given time (t). The value of the variables at any other time can be derived from these data and can be presented in the form of time trends. The solution of a simple model may be fairly easy, but the calculations involved in the solutions of more complex models may require the assistance of electronic computers.

Construction of the Model

The examples given below show how in the field of tuberculosis, a simplified epidemetric model can be constructed. It is not claimed that this model will be very realistic, but it is claimed that even the simplest of models may be profitable in so far as it represents an improved utilization of available data.

The basic steps in model-making are (1) definition of the mechanism, (2) estimation of the parameters, and (3) solution of the model.

(1) The Definition of the Mechanism —The definition of the mechanism of the present model is based on the following axioms:

(a) Tuberculosis is an infectious disease caused by the transmission of

* When defining the relationships one has to confine oneself to the most relevant factors. The possible influence of other factors which may sometimes be innumerable can however be pooled into one quantity, the properties of which have to be estimated. Models constructed this way are called stochastic models. In the present paper we will, for the sake of simplicity, deal with the so-called "exact" models. In these, all the "other" factors are supposed to even each other out. Furthermore, the exact models do not permit a direct measure of the accuracy of the estimated values.

tubercle bacilli from man to man; extrahuman sources of infection are disregarded.

(b) Tuberculosis is a benign infection in the sense that only a minor part of those infected develop disease (disabling tissue destruction).

(c) Only persons with tissue destruction (cases) can transmit infection to other persons.

In constructing the model it is not of principal importance how a "case" is defined. A case may be a person with certain types of x-ray lesions or a person excreting tubercle bacilli. The choice of definition will depend on the purpose for which the model is constructed and the nature of available data.

(d) Once infected, a person remains so for the rest of his life.

(e) Newborn are always free from infection.

On the basis of these axioms to which only few would object, the underlying mechanism in the spread of tuberculosis can be defined by a series of relationships and expressed in symbols. This process is demonstrated in Table 1.

Table 1 may be used for the formulation of the arithmetical relationships existing between the epidemiological quantities in the model. Line one gives the stock values (prevalences at the outset, time t). Lines two and three give the flow values (incidences) in and out, in the period between time t and t+1. By adding or substracting the flow values from the initial values the end results (stock values at time t+1, in line three) are computed.

The following simple relationships may be derived from Table 1 by adding vertically in each of the columns except for relationship (V) which is derived by adding horizontally in second row.

$$N_{t+1} = N_t + B_{t/t+1} - D_{N_{t/t+1}} - I_{t/t+1} \quad \text{(I)}$$

$$I_{t+1} = I_t + I_{t/t+1} + H_{t/t+1} - D_{I_{t/t+1}} - C_{t/t+1} \quad \text{(II)}$$

$$C_{t+1} = C_t + C_{t/t+1} - D_{C_{t/t+1}} - H_{t/t+1} \quad \text{(III)}$$

$$P_{t+1} = P_t + B_{t/t+1} - D_{t/t+1} \quad \text{(IV)}$$

$$D_{t/t+1} = D_{C_{t/t+1}} + D_{I_{t/t+1}} + D_{N_{t/t+1}} \quad \text{(V)}$$

Table 1—Definition of Symbols and Quantities in the Model

Stock Quantities (Prevalences at time t).	Noninfected (N_t)	Infected		Total Population (P_t)
		Noncases (I_t)	Cases (C_t)	
In flow (+) (Incidences)	Born [$B_{t/t+1}$]	Infected [$I_{t/t+1}$] Healed cases [$H_{t/t+1}$]	Converted to cases from infected noncases [$C_{t/t+1}$]	Born [$B_{t/t+1}$]
Outflow (−) (Incidences)	Death [$D_{N_{t/t+1}}$] Infected [$I_{t/t+1}$]	Death [$D_{I_{t/t+1}}$] Converted to cases [$C_{t/t+1}$]	Death [$D_{C_{t/t+1}}$] Healed cases [$H_{t/t+1}$]	Death [$D_{t/t+1}$]
Stock quantities [Prevalences at time (t+1)]	Noninfected [N_{t+1}]	Noncases. [I_{t+1}]	Cases [C_{t+1}]	Total population [P_{t+1}]

Equation (I) states that the number of noninfected at time (t+1) equals the number of noninfected at time t plus the number of newborn minus the number of noninfected persons who have died and persons who have become infected in the period under consideration.

Equation (II) states that the number of infected noncases at the end of the period is obtained by adding the number at the beginning to the number infected during the period, and deducting the number of infected persons who have died and who have been healed.

Equation (III) states that the number of cases at time (t+1) is obtained by adding to the number of cases at time t the number of new cases occurring during the period and deducting those who have died or have been healed.

Equation (IV) states that the population at time (t+1) is computed by adding the newborn in the period to the initial population, and subtracting the deaths.

Equation (V) states the simple fact that the total number of deaths is equal to the sum of deaths in each of the three subgroups of the population.

For the sake of simplification, the effect of migration is disregarded in this model.

Apart from the above relationships, another series of relationships, viz., functional relationships, exist between the various epidemiological variables. These relationships express how one quantity varies with variations in the others.

The functional relationships which appear to be of primary epidemiological interest are given below:

$$I_{t/t+1} = f_1(C_t) \qquad \text{(VI)}$$

$$C_{t/t+1} = f_2(I_t) \qquad \text{(VII)}$$

$$H_{t/t+1} = f_3(C_t) \qquad \text{(VIII)}$$

$$D_{N_{t/t+1}} = f_4(N_t) \qquad \text{(IX)}$$

$$D_{I_{t/t+1}} = f_5(I_t) \qquad \text{(X)}$$

$$D_{C_{t/t+1}} = f_6(C_t) \qquad \text{(XI)}$$

$$B_{t/t+1} = f_7(P_t) \qquad \text{(XII)}$$

The "f" states that there exists some relationship between the variable to the left and the variable in the parenthesis to the right.

Relationship (VI) says that the number becoming infected over a period of time depends upon the number of cases at the beginning of the period. In addition, one might think of a great many other factors which may affect the number of persons infected per case. Such factors are proximity of contact, composition of caseload (open or closed cases), and also number of uninfected available for new infection. For the sake of simplicity, these "other" factors have been disregarded in the construction of the present model. This simplification appears to be permissible if the period under study is so short that major environmental changes are unlikely to occur. The influence of availability of uninfected persons, for instance, may be disregarded as long as the proportion of uninfected is so high that lack of susceptible hosts is not a limiting factor in the spread of the tubercle bacilli.

The next relationship (VII) expresses that the number of new cases developing in a period depends upon the number of persons infected at the beginning of the period. This is based on the axiom that only infected persons become diseased.

The number of cases healed during the period is stated in relationship (VIII) to depend only upon the number of cases existing at the beginning of the period.

The three relationships (IX), (X), and (XI) define the mortality rates in the three subgroups of the population: the noninfected (N), the infected noncases (I), and the cases (C). Instead

of introducing the mortality rates, one could use the value, average length of life, the two of course being very closely related. Which of the two to select depends mostly on the data available.

Relationship (XII) states that the number of births is a function of the total population.

So far no mention has been made here about the form of the relationships. For the sake of simplification it will be assumed, in the present model, that the relationships in functions (VI)-(XII) are linear; this means, for instance, that the number of new persons infected is proportionate to the number of cases. This appears to be a reasonable assumption over a relatively short period in which environment and virulence of the microorganisms may be considered to remain the same. Over a long period this assumption would be increasingly unrealistic.

(2) Estimation of Parameters—The values of the parameters are estimated by actual observation of the sets of variables that enter the functional relationships (VI)-(XII). In function (VI), for instance, the "a" is the parameter to be estimated. This is done by observing C_t (number of cases at time t) and $I_{t/t+1}$ (number of new infections occurring over the period t to t+1) through some kind of epidemiological investigations in the area concerned. If several sets are available, the "a" may be estimated by means of a regression analysis; if only one set is available the best estimate of "a" will simply be $\frac{I_{t/t+1}}{C_t}$.

Considerable disagreement is likely to arise over the choice of parameters, especially when data have not been explicitly collected for the purpose of model-making and when therefore best estimates have to be made from whatever data are available. However, the advantage of the model approach lies in the fact that different sets of data may be used so that the trend may be shown under any assumptions which may be put forward.

(3) Solution of the Model—When the parameters have been estimated and the initial situation is given (in terms of prevalences), the model can be solved, i.e., each of the variables (total population, proportion infected, proportion uninfected, number of cases, and so on) can be expressed as functions of time and calculated for any point of time. How this is done is illustrated for a specific example in the Appendix.

Further Development and Refinement of the Model

The 12 relationships given above represent a dynamic model in its symbolic form. This simplified model should only be considered as an example of how the method works, and should not be taken as an exhaustive description of the epidemiological situation. Several refinements may be introduced into the mechanisms of this skeleton-model. This would make the model more realistic in the sense that more factors which are at play in the real world would be allowed to influence the outcome of the calculations. The type of refinements which one might wish to introduce depends to some extent on the purpose of the model-making. It would appear that for the purpose of judging impacts of various control programs, rather rough models may suffice as long as only the difference between one program and the other is sought. If, on the other hand, the object of the model is to determine the spontaneous trend, a considerable degree of refinement is probably called for. The following are some obvious refinements which could be included in the model, the first of which would seem to be needed even for the limited purpose of estimating relative impact of control programs.

(1) Distinction between open and

Table 2—Epidemiological Data Used for the Estimation of the Parameters

Prevalence of infection at time t (I_t)		25%	Own experience from studies in South India (not yet published). Frimodt-Møller estimates the prevalence to be 52 per cent in his material, but points out the difficulty in interpreting his findings because of the high prevalence of nonspecific sensitivity.
Prevalence of cases at time t (C_t)		1.5%	This is the prevalence of active and possibly active x-ray cases as given in Frimodt-Møller's report. This tallies with result from the National Sample Survey Report.[8]
Annual incidence of infection $\left(\frac{I_{t/t+1}}{P_t}\right)$		2.0%	Frimodt-Møller's estimates vary between 2-4 per cent. The low rate has been selected in conformity with the choice of low infection prevalence. With 1.5 per cent cases and 75 per cent non-infected, this means that each case infects one person per year (2% of 75% is 1.5%).
Annual incidence of new cases among infected $\left(\frac{C_{t/t+1}}{I_t}\right)$		0.85%	With 25 per cent infected and 2.1 per cent total incidence, this will be the incidence among the infected. Frimodt-Møller gives the incidence as 2.4 per cent. This was slightly reduced since his figure probably includes some lesions of non-tuberculous nature.
Annual rate of healing $\left(\frac{H_{t/t+1}}{C_t}\right)$		10%	Frimodt-Møller's findings indicate a healing rate of approximately 12 per cent based on changes in x-ray films from one round to the next. (Note: some treatment was given).
Annual death rates $\left(\frac{D_{t/t+1}}{P_t}\right)$	N	1.4%	Frimodt-Møller observes a total death rate of 1.6 per cent. Different age-composition of N and I and varying age-specific death rates makes similar death rate for the two groups justified.
	I	1.4%	
	C	7.0%	
Annual birth rate $\left(\frac{B_{t/t+1}}{P_t}\right)$		2.7%	Frimodt-Møller's direct observation. This is probably underestimated.

closed cases. If knowledge regarding the proportion of open and closed cases were available it would be profitable to take this proportion into account, since not only the infectivity but also the healing and the mortality may be different for the two categories of cases.

(2) Distinction between breakdown of infected into cases irrespective of the surrounding infectious cases and breakdown of infected into cases dependent on the surrounding infectious cases.

(3) Relating case incidence to the time elapsed after infection.

(4) Taking other changes by time into account. Other parameters such as healing and death rate may also be expressed as functions of time.

(5) Taking migration into account. It is evident that if more refinements are introduced into the model, the mathematical solution will become more complicated. The technical obstacles are, however, not unsurmountable even

when a good many more functions are included. The present model takes into account only the most basic relationships, since it was thought that such a presentation would best serve the purpose of maximum clarity.

Potentialities of the Model

In order to illustrate its potentialities, the model was solved by using data mainly derived from Frimodt-Møller's longitudinal survey in South India.[7] These data appear to be among the most complete which are available anywhere in the field of tuberculosis epidemiology. Even these data must, however, be taken with some reservations since inherent difficulties make certain estimates rather doubtful. Thus the high degree of nonspecific tuberculin sensitivity in the area greatly complicates the interpretation of tuberculin testing and makes it difficult to estimate how great a proportion of the population was actually infected with Mycobacterium tuberculosis. Furthermore, some treatment was given to cases detected during the survey, thus affecting the rates of infection, mortality, and healing in the survey population. Because of these shortcomings in the data and the rather embryonic stage of the present model, the following solutions of the model should not be taken to reflect the true epidemiological situation of tuberculosis in South India; they should merely be considered as examples of how the model is solved and what the potentialities of the method are.

Three examples of how the model is used are shown below:

1. Time Trend—This example is constructed by feeding a set of data largely based on the somewhat modified findings of Frimodt-Møller into the model and solving it over a period of 20 years. This is an illustration of the use of the

Table 3—Parameters Used in the Three Examples

Functional Relationships			Control Programs: Spontaneous Time Trend	Control Programs: One Mass Case Finding Program with Treatment for 1,000 out of 1,500 Cases	Control Programs: A Current BCG Program
			Ex. 1	Ex. 2	Ex. 3
$I_{t/t+1} = a \cdot C_t$		(VI)	1.0	1.0	0.5
$C_{t/t+1} = g \cdot I_t$		(VII)	0.0085	0.0085	0.0085
$H_{t/t+1} = e \cdot C_t$		(VIII)	0.1	0.1	0.1
$D_{N_{t/t+1}} = d_1 \cdot N_t$		(IX)	0.014	0.014	0.014
$D_{I_{t/t+1}} = d_2 \cdot I_t$		(X)	0.014	0.014	0.014
$D_{C_{t/t+1}} = d_3 \cdot C_t$		(XI)	0.07	0.07	0.07
$B_{t/t+1} = b \cdot P_t$		(XII)	0.027	0.027	0.027
Initial position	I_0		23,500	24,500	23,500
	N_0		75,000	75,000	75,000
	C_0		1,500	500	1,500

Table 4—Solutions of the Model for the Three Examples

General Solutions	Year t_0	t_5	t_{10}	t_{20}
Example 1				
$\frac{C_t}{P_t} = 0.01112 \cdot 1.0152^t + 0.00386 \cdot 0.772^t$	1.5	1.30	1.32	1.50
$\frac{I_t}{P_t} = 0.2552 \cdot 1.0152^t - 0.0202 \cdot 0.772^t$	23.5	27.0	29.6	34.6
Total prevalence of infection	25.0	28.3	30.9	36.1
Example 2				
$\frac{C_t}{P_t} = 0.00945 \cdot 1.0152^t - 0.00445 \cdot 0.772^t$	0.5	0.90	1.06	1.27
$\frac{I_t}{P_t} = 0.1962 \cdot 1.0152^t - 0.04878 \cdot 0.772^t$	24.5	21.2	23.2	26.5
Total prevalence of infection	25.0	22.1	24.3	27.8
Example 3				
$\frac{C_t}{P_t} = 0.01195 \cdot 0.992^t + 0.00305 \cdot 0.795^t$	1.5	1.24	1.13	1.02
$\frac{I_t}{P_t} = 0.25 \cdot 0.992^t - 0.01495 \cdot 0.795^t$	23.5	23.5	22.9	21.2
Total prevalence of infection	25.0	24.7	24.0	22.2

model to show the spontaneous time trend in a given epidemiological situation.

2. Interference of Case Finding and Treatment — This example illustrates how the model can be used to show the effect of a control program. The imaginary control program consists of a one-time mass case-finding and treatment scheme, in which we assume a very high operational efficiency, viz., that two-thirds of the cases are detected and successfully treated (permanently cured). The immediate effect of this control scheme is then fed into the time trend model of example 1, to show how such control measures would affect the trend of tuberculosis in the survey area.

3. Effect of a BCG Program—In this example it is assumed that a continuous BCG campaign succeeds in maintaining 70 per cent of all infected efficiently vaccinated in the area, and that BCG vaccine prevents 70 per cent of infection from taking place among the vaccinated. This results in a reduction of (0.7 x 0.7) approximately 50 per cent in the infection rate. This effect of the campaign (a 50 per cent reduction on incidence of infection) is fed into the time trend model from example 1 and solved over a period of 20 years.

The parameters used for the solution of the model were, as mentioned above, derived from Frimodt-Møller's survey in South India. The present authors have, however, made certain interpretations of the data which differ from those of Frimodt-Møller. The basis for the estimations is outlined in Table 2.

Table 3 gives the relationship (VI)-(XII) together with the parameter values for the three examples.

The solution was then obtained for

each of the three examples following the procedure shown in the Appendix, for example 1. The resulting equations and the calculated case prevalences at various points of time are shown in Table 4 and the case prevalence at various points of time are presented graphically in Figure 1.

It appears from Figure 1 that with the chosen dynamics of tuberculosis epidemiology, the trend of tuberculous disease may remain rather stable over the foreseeable future. Or to be even more careful—if the problem of tuberculosis decreases, it does so slowly and if it increases it cannot be with much speed. It also appears that the effect of even such a drastic effort as the discovery and cure of two-thirds of all cases in the area at one point of time does not markedly alter the trend of the disease since the case prevalence seems to approach the original level in the course of a few years. This apparently incongruous outcome of the model merely reflects a basic feature of tuberculosis epidemiology, viz., that new cases develop from the already existing pool of infected persons with a considerable time interval between infection and breakdown. A repeated BCG program of the suggested high efficacy, on the other hand, seems to be able to produce a real problem reduction which is apparent after five to ten years and which seems to increase with time as the pool of infected shrinks.

The epidemiological technics broadly suggested in this paper are a modest beginning and they are open to further development as more data and better understanding become available. It is the authors' hope that the examples given in this paper—and they are and should only be considered examples—may stimulate epidemiologists in other centers to attempt a similar approach so that clearer hypotheses may be formulated as soon as possible to be put at the test of time and experiments.

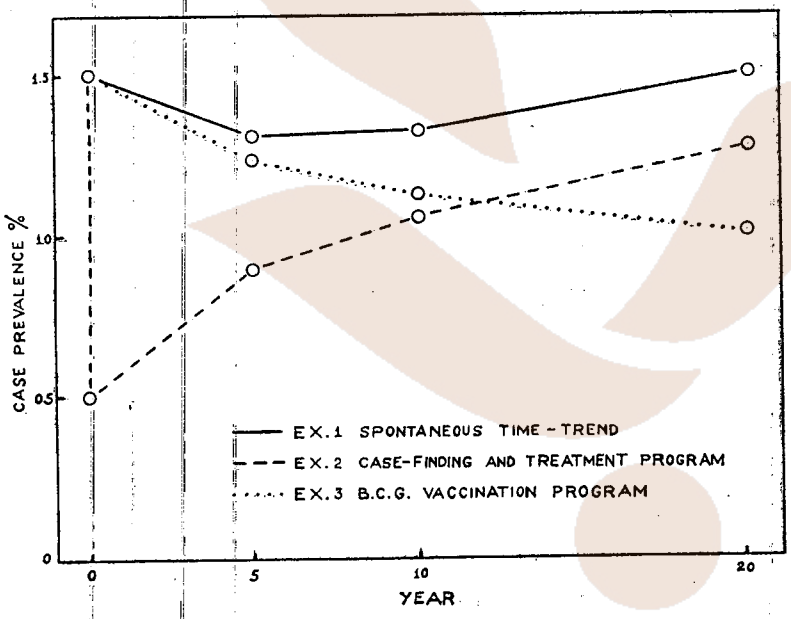

Figure 1—Examples of Model Solutions in Terms of Case Prevalence over 20 Years

Summary and Conclusions

The present paper gives examples of the use of epidemetric models which reflect the epidemiological trend of tuberculosis. These models were constructed by applying methods which have been developed and utilized in other social sciences.

Exact estimates of the various parameters entering the model must be available if realistic long-term results are to be achieved through model-making. The need for exact data regarding prevalences and incidences of infection and disease may necessitate longitudinal surveys in large random samples of population groups. It is, however, the present authors' firm opinion that it would be fruitful for almost any health department to combine their best available epidemiological knowledge in a system of relationships in order to quantify their concept of the situation. Such an exercise in mathematics would in any case serve to sharpen the epidemiologists' thinking and would lead them to appreciate what data they need most urgently.

The potentialities of the model are shown in three examples, illustrating how the model may help in predicting the trend of tuberculosis in a given situation, either spontaneously or under influence of specific control programs. These examples emphasize in particular that the model method may be profitable in evaluating specific control programs by reflecting their interference in the natural trend of tuberculosis in the control area.

REFERENCES

1. Hedrich, A. W. Monthly Estimates of the Child Population "Susceptible" to Measles, 1900-1931, Baltimore, Md. Am. J. Hyg. 17:613–636 (May), 1933.
2. Muench, Hugo. Catalytic Models in Epidemiology. Cambridge, Mass.: Harvard University Press, 1959.
3. Frost, W. H. How Much Control of Tuberculosis? A.J.P.H. 27:759, 1937.
4. Feldman, F. M. How Much Control of Tuberculosis 1937-1957-1977? Ibid. 47,10:1235–1241 (Oct.), 1957.
5. Palmer, C. E.; Shaw, L. W.; Comstock, G. W. Community Trials of BCG Vaccination. Am. Rev. Tuberc. 77:877, 1958.
6. Sutherland, I. An Estimation of the Scope for BCG Vaccination in Preventing Tuberculosis Among Those Aged 15-19 Years in England and Wales at the Present Time. Tubercle 40:413, 1959.
7. Frimodt-Moller, J. A. Community-Wide Tuberculosis Study in a South Indian Rural Population, 1950-1955. Bull. World Health Organ. 22:61, 1960.
8. Tuberculosis in India—A Sample Survey 1955-1958. New Delhi: Indian Council of Medical Research, 1959.

APPENDIX

Example of calculations of

$$\frac{C_t}{P_t} = F_1(t) \text{ and } \frac{I_t}{P_t} = F_2(t)$$

With the parameters given in Table 2, example 1, the equations (VI-XII) will take the following forms:

$$\begin{aligned}
I_{t/t+1} &= C_t && \cdot\cdot \text{ (VI)}\\
C_{t/t+1} &= 0.0085 \cdot I_t && \cdot\cdot \text{ (VII)}\\
H_{t/t+1} &= 0.1 \cdot C_t && \cdot\cdot \text{ (VIII)}\\
D_{N_{t/t+1}} &= 0.014 \cdot N_t && \cdot\cdot \text{ (IX)}\\
D_{I_{t/t+1}} &= 0.014 \cdot I_t && \cdot\cdot \text{ (X)}\\
D_{C_{t/t+1}} &= 0.07 \cdot C_t && \cdot\cdot \text{ (XI)}\\
B_{t/t+1} &= 0.027 \cdot P_t && \cdot\cdot \text{ (XII)}
\end{aligned}$$

The values are then inserted in equation (II):

$$I_{t+1} = I_t + I_{t/t+1} + H_{t/t+1} - D_{I_{t/t+1}} - C_{t/t+1}$$

$$I_{t+1} = I_t + C_t + 0.1 \cdot C_t - 0.014 \cdot I_t - 0.0085 \cdot I_t$$

$$C_t = \frac{I_{t+1} - 0.978 \cdot I_t}{1.1} \cdot\cdot \text{ (A)}$$

APPENDIX (Continued)

and in equation (III)

$$C_{t+1} = C_t + C_{t/t+1} - Dc_{t/t+1} - H_{t/t+1}$$
$$C_{t+1} = C_t + 0.0085 \cdot I_t - 0.07 \cdot C_t - 0.1 \cdot C_t$$
$$C_{t+1} - 0.83 \cdot C_t = 0.0085 \cdot I_t \quad \cdots \quad (B)$$

From these two equations (A) and (B), both C_t and I_t can be expressed implicitly as functions of the time (t):

$$C_{t+2} - 1.808 \cdot C_{t+1} + 0.802 \cdot C_t = 0 \quad \cdots \quad (C)$$

and

$$I_{t+2} - 1.808 \cdot I_{t+1} + 0.802 \cdot I_t = 0 \quad \cdots \quad (D)$$

and with the initial position given, by usual difference equation technique, explicitly as functions of the time.

$$C_t = 1112 \cdot 1.027^t + 388 \cdot 0.781^t$$
$$I_t = 25525 \cdot 1.027^t - 2025 \cdot 0.781^t$$

By inserting the values of equations (VI), (IX) and (XII) in equation (I), one gets:

$$N_{t+1} = N_t + B_{t/t+1} - D_{N\,t/t+1} - I_{t/t+1}$$
$$N_{t+1} = N_t + 0.027 \cdot P_t - 0.014 \cdot N_t - C_t$$

As $N_t = P_t - C_t - I_t$ and from equation (B)

$$I_t = \frac{C_{t+1} - 0.83 \cdot C_t}{0.0085}$$ one gets:

$$0.0085 \cdot P_{t+1} - 0.0086 \cdot P_t - C_{t+2} + 1.808 \cdot C_{t+1} - 0.802 \cdot C_t = 0 \quad (E)$$

Equation (C) above indicates, however, that the right part of the equation (E) = 0, giving the simple solution:

$$0.0085 \cdot P_{t+1} - 0.0086 \cdot P_t = 0$$

or

$$P_t = P_0 \cdot 1.0118^t = 100000 \cdot 1.0118^t$$

The prevalence of cases can then be expressed as a function of time.

$$F_1(t) = \frac{C_t}{P_t} = \frac{1112 \cdot 1.027^t + 388 \cdot 0.781^t}{100000 \cdot 1.0118^t} = 0.01112 \cdot 1.0152^t + 0.0038 \cdot 0.772^t$$

Similar for the prevalence of infected (noncases)

$$F_2(t) = \frac{I_t}{P_t} = \frac{25525 \cdot 1.027^t - 2025 \cdot 0.781^t}{100000 \cdot 1.0118^t} = 0.2552 \cdot 1.0152^t - 0.0202 \cdot 0.772^t$$

The sum of these two expressions $\left(\frac{C_t}{P_t} + \frac{I_t}{P_t}\right)$ gives the prevalence of infection as a function of time.

At the time of this study, Mr. Waaler was WHO senior statistician, Dr. Geser was WHO epidemiologist, and Mr. Andersen was senior WHO officer, World Health Organization, National Tuberculosis Institute, Bangalore, India.

The views expressed in this paper are those of the authors and are not necessarily the views of the World Health Organization.