INTERPRETATION OF TUBERCULOSIS INFECTION AGE CURVES

JØRGEN NYBOE, cand. act.
Statistician, WHO Tuberculosis Research Office, Copenhagen, Denmark

SYNOPSIS

The percentage of tuberculin reactors, by age, is sometimes used in the epidemiology of tuberculosis as one measure of the tuberculosis problem in a population. Whereas such data are useful for determining the load of infection in communities, it is shown in this paper that only in very few instances do they lend themselves to a quantitative determination of the risk of infection. Such data are, therefore, of limited epidemiological value, and the author suggests other types of data which could give more useful information.

The results of tuberculin-testing obtained in sixteen mass BCG campaigns, from 1948 to 1951, have been used to illustrate the problems under discussion.

One characteristic of the tuberculous infection is that once a person has become infected the fact can be demonstrated for many years to come by means of the tuberculin test. The tuberculin test has, therefore, won recognition as an important epidemiological tool in tuberculosis control work, and tuberculin-testing surveys are being carried out today on a large scale in most parts of the world.

These surveys are usually conducted by testing the population under study, or a representative sample of it, at a certain point in time, whereby an estimate is obtained of the proportion of persons infected with tuberculosis. This gives some indication of the risk of infection and thus a measure of the tuberculosis problem in the community.

Whereas the principle of a tuberculin survey of this kind is very simple, the problem of precisely what information can be derived from the data collected, and how the data can be used for a quantitative evaluation, is less simple and has scarcely been explored; most authors have been content merely to present the data obtained and compare the frequencies of infected persons in the different groups. The purpose of the present paper is to analyse the data on frequencies of infected persons obtained by tuberculin-testing a population once.

The first obvious point in the analysis is that these data must be broken down by age. As the skin sensitivity to tuberculin that results from infection
is a chronic manifestation, the results of tuberculin-testing will reflect the accumulated effect of exposure from birth until the time of testing, and age, or the length of exposure, is therefore a decisive factor. Data on frequencies of tuberculosis-infected persons by age are usually presented graphically as curves and will here be termed, for brevity, tuberculosis infection age curves.

As a second point in the analysis of the curves, it is emphasized that the risk of infection probably changes with calendar time and that the likelihood of such changes seriously limits the value of the curves for determining the risk of infection. Though this limitation has been recognized in a few instances,\textsuperscript{3,4} it has been ignored as well;\textsuperscript{1} and an analysis of the infection age curve in relation to the infection risk is therefore called for.

Despite the limitations described, the infection age curves may under certain conditions be used to compare the level of infection risk among different populations, and a method for comparison is suggested in this paper. Finally, survey methods designed to yield more useful epidemiological information are briefly discussed.

**Examples of Tuberculosis Infection Age Curves**

To illustrate the problems under discussion use has been made of tuberculin-test results from mass BCG campaigns conducted under the auspices of the International Tuberculosis Campaign (ITC)\textsuperscript{a} in 1948-51.\textsuperscript{6}

These data have certain features that make them well suited for the purpose. First, they are derived from many different countries and territories (Austria, Czechoslovakia, Greece, Malta, Poland and Yugoslavia in Europe; Algeria, Egypt, Morocco, Tangier and Tunisia in North Africa; Israel, Lebanon and Syria, as well as the Palestine refugees, in the Middle East; and Ecuador in South America). Thus they can be expected to provide a wide variety of infection age curves. Secondly, the testing methods employed were kept uniform, as far as possible, throughout the campaigns, so that the data should be comparable from country to country. Young children were usually tested with the Moro patch-test. An intradermal (Mantoux) test with 5 TU or 10 TU was given to older children and adults, sometimes as a single test, sometimes subsequent to a weaker test.\textsuperscript{b} With a few exceptions the tuberculin used was prepared by one laboratory. The method of reading the test and the criteria for positive reactions were also the same in all the campaigns. Finally, the proportion of the populations attending was very high in these campaigns, and test results are available for more than 20 million persons. Sampling errors are therefore small, even when the data are broken down by age and district.

\textsuperscript{a} A joint enterprise of UNICEF, the Danish Red Cross, the Norwegian Relief for Europe and the Swedish Red Cross

\textsuperscript{b} In one country, Malta, the Adrenalin-Pirquet test was used for all ages.
When data on frequencies of tuberculin-positives are to be analysed, the first problem to consider is, of course, whether those classified as positives are in fact those infected with tuberculosis. This problem will not be discussed in detail here, but it should be mentioned that investigations made in the last ten years have shown that in many parts of the world an intra-
FIG. 2. OBSERVED TUBERCULOSIS INFECTION AGE CURVES FOR AREAS CHOSEN TO REPRESENT VARIOUS LEVELS OF POSITIVITY IN SIXTEEN ITC CAMPAIGNS

AUSTRIA CZECHOSLOVAKIA GREECE

MALTA POLAND YUGOSLAVIA

ALGERIA EGYPT MOROCCO & TANGIER

TUNISIA ISRAEL LEBANON

PALESTINE REFUGEES SYRIA ECUADOR

The age at which the test was changed from Mora to Montoux is indicated by a broken line.
dermal test with a low dose of tuberculin—5 TU or 10 TU—will separate the population into two groups, one with large-sized “positive” reactions, the presumably infected, and one with zero or small “negative” reactions, the presumably uninfected. In some areas, mainly in the tropics, the test, however, fails to effect a separation, probably because of widespread non-specific sensitivity to tuberculin. So far as is known, non-specific sensitivity is not a problem in any of the countries for which test data are reported here; and it is assumed in the following that the results of the Mantoux test give fairly good estimates of the percentages of infected. The Moro test, however, seems to have been somewhat weaker than the Mantoux test in the ages 6-12 years.

Fig. 1 gives infection age curves for ten countries where the campaign covered the entire, or practically the entire, territory. The curves are arranged in three groups: the European area, the North African area, and Ecuador and Israel. The curves all show a rise with increasing age, but they differ considerably in shape and level. The European curves are more or less distorted by a sudden increase between the ages of 12 and 13 years where the test was shifted from Moro to Mantoux, but there is a general pattern, common to all five, of relatively low percentages of positives in the youngest age-groups and an accelerating increase in the percentages with increasing age. The level of the curves varies, however: the highest percentages of positives were found in Poland; lower percentages were observed in Yugoslavia and Czechoslovakia, and the lowest—and nearly the same—in Austria and Greece.

The infection age curves for Tunisia, Egypt and Algeria closely resemble each other (the dip at ages 8-12 in the curve for Egypt is probably due to the inefficiency of the Moro test in this age-range). Their pattern, however, clearly differs from that of the European curves: in the North African countries the percentages of positives rise more sharply in the young age-groups and tend to level off with increasing age.

The curve for Ecuador resembles the European curves. The curve for Israel, on the other hand, is very irregular and shows no similarity to either the European or the North African pattern. The irregularity may in part be explained by the fact that the Israeli population tested included a high proportion of immigrants newly arrived.

Further examples of infection age curves are shown in Fig. 2. These curves are for large administrative areas representing high, medium and low levels of positivity found in sixteen ITC campaigns. Study of the European curves in Fig. 2 shows that the curves for areas with low levels of positivity all increase in steepness with age. A different pattern is found for the curves for the high-level areas: straight-line curves characterize the higher levels of positivity in Vojvodina, Yugoslavia, and Szczecin, Poland.

---

a The curves for Tunisia and Algeria where the Moro test was only used up to the age of 6 years show no such interruption.
and the curve for Łódz city, where the percentages are highest, is steepest in childhood and then levels off with increasing age. On the other hand, none of the North African curves, not even those at the lowest levels, show increasing steepness with age.

**FIG. 3. DIFFERENCES BETWEEN TUBERCULOSIS INFECTION AGE CURVES FOR MALES AND FEMALES IN THE TEN COUNTRIES SHOWN IN FIG. 1**

The percentages of positives in Fig. 1 and 2 are averages for males and females. The differences between the percentages for males and for females for the curves given in Fig. 1 are shown in Fig. 3. Little or no difference is seen among young children, but after a certain age the percentages of positives among males in all countries begin to exceed those among females, and, the difference tends to increase with age. The age at which males show more positive reactions than females varies in the different areas: in the European countries it is around 16 years; in Africa much earlier—about 10 years; in Ecuador and Israel around 12 years. However, the sex differentials are on the whole small, and averaged percentages have been used throughout this paper.

**Theoretical Analysis of Infection Age Curves**

The interpretation of the curves shown above poses several problems, some related to the levels of the curves, others to the shape of the curves.
The curve for Poland, for example, runs at a higher level than that for Austria, indicating that the exposure to infection is—or rather has been—higher. But how much higher? And what is the significance of the shape of the infection age curve? Does the shape permit us to conclude that the risk of infection is higher at some ages than at others? The difference in shape between the North African and the European curves clearly indicates that the risk of infection has been different in the two regions, but is it possible to find out wherein the difference lies?

To answer questions of this kind the relation between the tuberculosis infection age curve and the annual infection rate must be studied. The annual infection rate—the proportion of uninfected persons who become infected during the year—is the most direct measure of the risk of infection in a population group. By means of annual infection rates the risk of infection in different age-groups, different social groups, different countries, etc. can be compared. There is evidently some relation between the annual infection rate and the infection age curve, for example, the higher the rate the higher the level of the curve, and in order to study this relation we shall in the following make different assumptions about the rate and compare the resulting curves. Through this indirect approach, it will be possible to determine whether or when annual infection rates can be derived from observed infection age curves.

Let us begin with the most simple assumption—that the annual infection rate remains the same from calendar year to calendar year—and let us further assume that the infection rate is the same for all ages. If, say, 5% of tuberculin-negative persons in all ages become tuberculin-positive during one year, then the proportion of negatives among persons 1 year old will be 0.95, among those 2 years old (0.95)² ... among those n years old (0.95)ⁿ, and the proportion of positives at n years of age will be 1−(0.95)ⁿ. The infection age curve that would correspond to this infection rate is given as curve I in Fig. 4. The curve is steepest in early childhood and levels off steadily with increasing age.

If the annual infection rate changes with age (but still remains the same from calendar year to calendar year), other types of infection age curves may result. If, for example, the rate increases linearly with age as follows: 2.0% for persons 0-1 year old, 2.6% for those 1-2 years old, 3.2% for those 2-3 years old ... 8.0% for those 10-11 years old, etc., then the proportion of positives at n years of age will be 1−0.980×0.974×0.968× ... (0.980−0.006(n−1)). The infection age curve constructed for these rates is shown as curve II_a in Fig. 4. The curve is nearly a straight line from the age of 3 onwards, and it intersects curve I. Doubling the infection rates of II_a will produce the convex infection age curve II_b, and halving the rates will result in the concave curve II_c. Although the infection rates determining

---

*a Formulae dealing with this relation are given in Annex 1 (see page 335).*
curves $\Pi_a$, $\Pi_b$ and $\Pi_c$ increase by age in the same proportion, the curves differ in shape as the infection rates differ in level.

When the rates of infection remain constant from calendar year to calendar year—irrespective of whether the rates are the same for all ages or vary with age—the percentage of positives at any one age is uniquely determined by the prevailing infection rates. Conversely, the annual infection rate for each age can be computed from the curves. The computation (which follows from the principle of the construction of the infection age curve) consists in subtracting the percentage of negatives at one age from the percentage of negatives at the preceding age and dividing this difference by the percentage at the younger age. For example, in curve $\Pi_a$, the percentage of negatives at age 11 is 56.75 and at age 10 is 61.69; the infection rate for persons between 10 and 11 years of age is therefore $\frac{4.94}{61.69} = 0.080$ or 8.0%, as assumed.

But let us now assume that the rates of infection change from calendar year to calendar year. The exposure of persons born in different years will consequently be different, and the increase in the percentage of positives from one age to the next will not reflect the prevailing infection rate for the age-interval. This is illustrated below in a constructed example (see also Annex 2, page 336): the annual infection rate as a function of age ($x$) and calendar time ($t$) is denoted by $\rho(x, t)$ and for simplicity it is assumed that this function consists of two factors, an age component $f(x)$ and a time component $g(t)$, i.e., $\rho(x, t) = f(x)g(t)$. Let us next assume that in an area $A$ the infection rate as a function of age is given by $f^A(x) = 0.02 + 0.002x$ and that $g^A(t) = 1$ and has been so for a long time. The infection age curve
TUBERCULOSIS INFECTION AGE CURVES

will then be as curve \( A^0 \) in Fig. 5. In another area, B, the rates increase with age by the same proportionate increment as do those in area A, but the level of the rates is twice as high, i.e., \( f^B(x) = f^A(x) \), but \( g^B(t) = 2 \). When these rates have prevailed for many years the infection age curve will look like \( B^0 \). Suppose now that at a certain point in time, \( t_0 \), conditions begin to change in both areas; in area A the infection rates for all ages increase linearly to the double during 12 years; in B the rates decrease to half during 12 years. The trends by calendar time, i.e., the functions \( g^A(t) \) and \( g^B(t) \), are shown in Fig. 6. After 6 years the infection rates in the two areas will be the same \( (g^A(t_0 + 6) = g^B(t_0 + 6) = 1.5) \), but the infection age curves that would obtain at this time are those shown as \( A^0 \) and \( B^0 \) in Fig. 5. Although the infection rates are the same in the two areas at this point in time, the curves are far from identical: the curve for area A is still much lower than that for area B, except in early childhood. Both curves, in fact, closely resemble the curves found before the change began (\( A^0, B^0 \)), because in all but the youngest persons most of the exposure to infection occurred under the previous conditions.

If the rates continue to increase in area A and to decrease in area B, after 6 more years the rate of infection in area A is twice that in area B. The curve for area A at this time, \( A^{12} \), shows higher percentages of positives than the corresponding curve for area B, \( B^{12} \), up to age 14, but after that age it shows lower percentages despite the fact that the infection rates in area A are now higher than in area B.

**FIG. 6. HYPOTHETICAL TRENDS IN LEVELS OF INFECTION RATES BY CALENDAR TIME IN AREAS A AND B**

Furthermore, although the infection rates in area A after 12 years are the same as the rates in area B were before any change began, the curve \( A^{12} \) shows considerably lower percentages than the curve \( B^0 \) for all but the very youngest ages. Clearly, if annual infection rates were to be computed from the two curves \( A^{12} \) and \( B^0 \) by the method described above (see page 326) the results would not be the same. The correct rates could be computed from \( B^0 \) because the rates in area B at time \( t_0 \) have prevailed for a long time and therefore determine the curve. Annual infection rates in area A after
12 years cannot be obtained from the curve \( A^{13} \), because the curve reflects changing infection rates. If the method nevertheless is applied to this curve, it will be found that the computed rates for very young children (who have only been exposed to the recent rates) are approximately correct, but for older persons the error is considerable: in the age-group 15-19 years the rates computed from \( A^{12} \) are less than 60% of the correct values. Similarly, infection rates computed from \( B^{13} \) for the age-group 15-19 are about 170% of the correct values.

The assumptions made in this example of the degree of change in infection rates with calendar time are arbitrary and have been made just to illustrate the point. The magnitude of the changes, however, would not seem to be exaggerated; data reported by Myers et al.\(^8\) from Minnesota, for instance, indicate a much more rapid decrease in infection than was assumed in the example.

From the discussion of constructed infection age curves it is clear that observed curves reflect not only the present but also past experience of infection. It follows that if the data available consist of infection age curves found at one point in time—as is the case for the ITC campaign data—it is not possible to derive the infection rates prevailing at the time. But it should be pointed out that if infection age curves for earlier periods were available, reliable estimates of the current infection rates as well as an estimate of the trend in the rates with calendar time could be obtained. Obviously if a curve found 10 years earlier shows the same percentages as the present curve, infection may be assumed unchanged with calendar time and the rates for individual ages may accordingly be computed. And even if the past curve were different from the present—for example, if it were known not only that the present curve looked like \( A^{12} \) but also that the curve of 6 years ago looked like \( A^{6} \)—there would be a basis for obtaining valid estimates of the present rates and of the increase in the rates by calendar time.

**Comparison of Infection Rates by Infection Age Curves**

Although a single infection age curve cannot give an adequate picture of the infection rates in the population, there is still the possibility that contemporary infection age curves for different populations might be used for comparison of infection rates. Conclusions cannot be drawn from such comparisons, however, when the curves differ as much as they do between, for instance, North Africa and Europe. In considering, for example, the curves for Tunisia and Czechoslovakia (reproduced from Fig. 1 in Fig. 7), several explanations of the differences in shape might be given: the rates for younger persons may be higher and those for older persons lower in Tunisia than in Czechoslovakia; it may also be that there is a larger increase or a smaller decrease in the rates with calendar time in
Tuberculosis infection age curves for Tunisia and Czechoslovakia

Tunisia than in Czechoslovakia; or, in fact, the explanation may be many combinations of different age and time trends. Unless other data are available there is no basis for determining which of the many possible explanations is the correct one, and we can therefore make no comparative statement whatsoever about the infection rates in the two countries.

It follows that comparisons of infection age curves can only be fruitful if some similarity can be demonstrated between the annual infection rates. As a working hypothesis the assumption could be made that the infection rates have differed only in level, that changes in the rates by age and calendar time have been proportionately the same. If the infection rates in the different areas are denoted by \( \rho_1(x,t) \), \( \rho_2(x,t) \), etc., this relation would be expressed mathematically as follows:

\[
\begin{align*}
\rho_1(x,t) &= k_1 \rho(x,t) \\
\rho_2(x,t) &= k_2 \rho(x,t) \\
\text{etc.}
\end{align*}
\]

where \( k_1, k_2, \ldots \) are constants representing the various levels of the infection rate.

Offhand it would not be unreasonable to assume a basic similarity of infection rates within the European countries covered by the ITC programme: the socio-economic development and the way of life have been generally parallel, and all have been subject to the severe deprivations of the Second World War. An assumption of similarity might also be made for the four countries in the North African area. But we have already seen from the infection age curves that the infection rates for the combined European—North African region could not be similar.

In the following it will be investigated whether the simple mathematical relation given above holds true for the European area. The European
countries have been chosen because the percentages of positives vary widely and the validity of the relations can thus be studied over a wide range.

First, an average was made of the infection age curves for Austria, Czechoslovakia, Greece, Poland and Yugoslavia. By comparing an observed curve with the average curve for the whole area, an estimate of the level of infection \( k \) was computed, using the method of least squares, and by applying the \( k \) value to the average curve the infection age curve that would be found if our assumption were correct was then computed. The formulae and an example of the computations are given in Annex 3 (see page 337).

The computed curves for the five European countries are compared with the observed curves in Fig. 8. The agreement between observed and computed values is good, and although there are some systematic deviations the hypothesis that the infection rates have differed only with respect to level seems to hold rather well.

**FIG. 8. COMPARISON OF OBSERVED (-) AND COMPUTED (++) TUBERCULOSIS INFECTION AGE CURVES IN FIVE EUROPEAN COUNTRIES**

![Graph showing comparison of observed and computed tuberculosis infection age curves in five European countries.](image)

To study the agreement further, computations were made with the curves representing areas with high, medium and low levels of positivity in the European countries (Fig. 2); the computed curves are compared with the observed curves in Fig. 9. The differences between observed and computed values are somewhat larger than in Fig. 8, but the agreement is still quite satisfactory.

Fig. 8 and 9 show that our hypothesis about annual infection rates in the European area holds true with good approximation. It is therefore possible, even though the actual infection rates as functions of age and time are not known, to compare the level of the rates in these countries. Thus,
FIG. 9. COMPARISON OF OBSERVED (−) AND COMPUTED (++) TUBERCULOSIS INFECTION AGE CURVES FOR THE EUROPEAN AREAS SHOWN IN FIG. 2

The computed values (+) are not connected to form curves.

as the level factor, or $k$ value, for Poland is 1.58 and for Greece 0.597, the level of infection rates in Poland has been $\frac{1.58}{0.597} = 2.6$ times higher than in Greece, 2.6 times higher than in Austria, 1.7 times higher than in Czecho-
FIG. 10. GEOGRAPHICAL DISTRIBUTION OF LEVELS OF TUBERCULOSIS INFECTION RATES IN FIVE EUROPEAN COUNTRIES (BASED ON PERCENTAGES OF TUBERCULIN-POSITIVES AT THE AGE OF 10 YEARS)

PERCENTAGE OF POSITIVES

9.7 or less
9.8 - 14.2
14.3 - 20.5
20.6 - 29.2
29.3 - 40.4
40.5 - 54.0
54.1 or more
Data not available
slovakia and 1.2 times higher than in Yugoslavia. A level ratio thus computed would seem to be a valid yardstick for comparing the risk of infection in different population groups and, hence, of considerable importance in the epidemiology of tuberculosis.

An important consequence of a similarity of infection rates is that the percentages of positives at any single age can be used, as well as the $k$ values, as indices of infection risk—a more convenient procedure when many infection age curves are to be compared. But unless a similarity in infection rates can be demonstrated, the percentage of positives at one age can be misleading as an index. This is well illustrated by the curves for Czechoslovakia and Tunisia shown in Fig. 7, where a comparison by means of the percentages of positives will depend on the age selected: up to the age of 16 years the percentage of positives is higher in Tunisia than in Czechoslovakia, while at ages over 16 the opposite is true.

As the infection rates in the five European countries would seem to differ only with respect to level, it is possible to give a more detailed picture of the geographic variation of infection rates within this area. The map in Fig. 10 shows the level of the infection rates by administrative district (nearly 1600 in all) in each of the five countries. The percentages of positives at the age of 10 years, smoothed by the three-point moving-average method, have been used to characterize the levels of the rates. In order to give a visual impression, different shading has been used to indicate different levels of percentages. The limits for percentages with the same shading have been determined in such a way that the increase from the lower limit to the higher corresponds to an increase of 50% in the infection rate.

The infection rates show definite geographic trends in level, and there is a smooth continuity across country borders. (Only the western part of the Polish-Czechoslovakian border shows a discontinuity, and this may be due to the huge displacement of populations immediately after the war.) The percentages of positives on the map range from 5.7 in Laconia Department, Peloponnesus, to 63.1 in Senj District, Croatia, corresponding to a difference of the order of 1:17 in the infection rates.

Discussion

The percentage of positives at any given age is the end-result of the exposure to tuberculous infection throughout the lifetime of that age-group. This accumulated result cannot reveal the development of the risk of infection—how it varied through childhood, the school years, etc. Consequently, prevailing infection rates cannot be derived from the infection age curve, but it has been shown that levels of infection rates can in some instances be compared by means of the curves. Where the development of infection has apparently been similar, infection age curves can be used to
compare the levels of infection by area, density of population, socio-
economic group, etc. In areas where no other tuberculosis statistics are
available such a comparison does provide useful information about the
relative magnitude of the tuberculosis problem.

It must be admitted, however, that the comparison of levels has a limited
field of application. The data given in this paper, though informative for
countries within Europe, could not be used to make comparisons between
North Africa and Europe. The value of infection age curves obtained at a
single point in time will always be limited because they do not permit
deduction of infection rates—neither the present rates nor the trend. It
might, therefore, be more profitable to shift the emphasis in future work
from single tuberculin surveys to repeat surveys. However, repeat surveys
are not very useful if, as is often the case, those classified as non-reactors
in the first survey are vaccinated with BCG. It is not possible to ascertain
whether or not a vaccinated person has been infected between the initial
and the repeat survey, as with the present method of tuberculin-testing we
cannot distinguish between naturally acquired and BCG-induced tuberculin
allergy. Studies have been made indicating that a more specific tuberculin
test may be developed, but until such time the combining of tuberculin
surveys with BCG vaccination will seriously limit the possibility of deriving
infection rates by repeated surveys.

Repeated tuberculin surveys might be carried out in several ways. One simple scheme would be to test a representative group of persons of
the same age at regular intervals. Although this would not provide data on
infection rates, an indication would be obtained of the trends of the rates
and of the effect of tuberculosis control measures. To make the assessment
of the trend as precise as possible, an age-group with a short history of
exposure—for example, young children entering school—should be selected.

A more ambitious scheme would be a direct observation of infection
rates by testing a representative sample of the population and retesting the
same persons after a relatively short interval, perhaps one or two years.
Several problèmes would be met with in such a scheme, however: the tracing
and identification of persons at the retest is difficult even in the most well-
organized communities; and the analysis would be complicated by the
percentage of persons who show intermediate-sized doubtful reactions that
cannot be classified as either positive or negative. It would, therefore, be
more practical and equally useful not to attempt a follow-up but simply
to test representative samples of the population with an interval of, say,
5-6 years. This scheme would allow derivation of infection rates by age
and calendar time, and is simple and practical enough to be applicable
in nearly every population group.
Annex 1

FORMULAE FOR RELATION BETWEEN INFECTION RATES AND INFECTION AGE CURVES

For simplicity it is first assumed that the functions under investigation do not change with calendar time. Let \( \mu_x \) be the force of mortality in the population, i.e., the probability of a person of age \( x \) dying in the interval from \( x \) to \( x+dx \) is \( \mu_x dx \). The function \( \mu_x \) is related to the survival function \( l_x \) by the formula \( dl_x = -l_x \mu_x dx \), or \( l_x = l_0 e^{-\int_0^x \mu_x dr} \), \( l_0 \) being an arbitrary constant.

Further, let the function \( \lambda_x \) denote the force of mortality among tuberculin-negative persons, i.e., the probability of a negative person of age \( x \) dying between \( x \) and \( x+dx \) is \( \lambda_x dx \); and the function \( p_x \) the force of tuberculous infection, i.e., the probability of a tuberculin-negative person of age \( x \) becoming positive between \( x \) and \( x+dx \) is \( p_x dx \).

If \( n_x \) is the number of negatives at age \( x \), and assuming that reversions from positive to negative do not take place, we have:

\[
dn_x = -n_x (\lambda_x + p_x) dx, \text{ or } n_x = n_0 e^{-\int_0^x (\lambda_x + p_x) dr}
\]

where \( n_0 = l_0 \), since everyone is born negative.

The proportion of negatives at age \( x \), \( q_x \), will then be:

\[
q_x = \frac{n_x}{l_x} = e^{-\int_0^x (\lambda_x + p_x - \mu_x) dr}
\]

If the mortality among tuberculin-negative persons can be supposed nearly the same as that of the general population (i.e., if \( \lambda_x \) approximately equals \( \mu_x \)), the following formulae hold true with good approximation:

\[
(1) \quad q_x = e^{-\int_0^x \rho_x dr}
\]

\[
(2) \quad \rho_x = -\frac{1}{q_x} \frac{dq_x}{dx}
\]

When working with the annual infection rate, \( r_x \), instead of the force of infection, \( \rho_x \)* (the two functions are related by the formula: \( 1 - r_x = e^{-\int_x^{x+1} \rho_x dr} \), we find corresponding to formulae (1) and (2):

\[
(3) \quad q_x = (1 - r_x) (1 - r_1) \ldots (1 - r_{x-1}), \text{ and }
\]

\[
(4) \quad r_x = \frac{1}{q_x} \Delta q_x = \frac{q_x - q_x + 1}{q_x}
\]

* The distinction between the continuous and the discontinuous function has not been maintained in the text.
Suppose now that the functions under investigation change from calendar year to calendar year. The symbols \( p(x, t), q(x, t) \), etc., where \( x \) again denotes age and \( t \) is calendar time, will then be used. Formulae (1) and (2) will then be expressed as:

\[
(5) \quad q(x, t) = e^{-\int_{0}^{x} \rho(\tau, t - x + \tau) \, d\tau}
\]

\[
(6) \quad p(x, t) = -\frac{1}{q(x, t)} \left( \frac{\delta q(x, t)}{\delta x} + \frac{\delta q(x, t)}{\delta t} \right)
\]

Whereas an estimate of \( \frac{\delta q}{\delta x} \) can always be derived by numerical differentiation of the infection age curve, an estimate of \( \frac{\delta q}{\delta t} \) can only be obtained if infection age curves are available for two or more points in time.

Annex 2

AN EXAMPLE OF CHANGES IN INFECTION RATES BY CALENDAR TIME

Assuming that the force of infection can be broken into two components, an age factor and a time factor, i.e., \( p(x, t) = f(x)g(t) \), formula (5) in Annex 1 can be expressed as

\[
-\log_{e} q(x, t) = \int_{0}^{x} f(\tau)g(t-x+\tau) \, d\tau
\]

Let us put \( f(x) = a+bx \), and

\[
g(t) = \begin{cases} 
  k & \text{for } t \leq t_0 \\
  k+(t-t_0) & \text{for } t > t_0
\end{cases}
\]

(i.e., \( g(t) \) is constant until the time \( t_0 \) and then increases or decreases linearly)

(1) When \( t \leq t_0 \), the proportion of negatives at age \( x \) can be expressed as:

\[
-\log_{e} q(x, t) = -\log_{e} q(x, t_0) = \int_{0}^{x} (a+br) \, kdt = k(ax + \frac{bx^2}{2})
\]

(2) When \( t > t_0 \), two possibilities must be distinguished:

(i) that persons at age \( x \) are born after \( t_0 \);

(ii) that they are born before \( t_0 \).

(i) \( t-x > t_0 \)

\[
-\log_{e} q(x, t) = \int_{0}^{x} (a+br) \, (k+(t-x+\tau-t_0)) \, d\tau
\]

\[
= \int_{0}^{x} (a+br) \, kdt + \int_{0}^{x} k(a+br) \, (t-x+\tau-t_0) \, d\tau
\]

\[
= -\log_{e} q(x, t_0) + \frac{lx}{6} (6a(t-t_0) - 3ax + 3bx(t-t_0) - bx^2)
\]

or if we set \( t-t_0 = n \)

\[
-\log_{e} q(x, t_0+n) = -\log_{e} q(x, t_0) + \frac{lx}{6} (6an - 3ax + 3bzx - bx^2)
\]
(ii) \( t-x \leq t_0 \), that is, \( x \geq n \)

\[
-\log_q q(x, t_0+n) = \int_0^x (a+br) \, k \, dr + \int_{x-n}^x (a+br) (k+n-x+r) \, dr
\]

\[
= \int_0^x (a+br) \, k \, dr + \int_{x-n}^x (a+br) (n-x+r) \, dr
\]

\[
= -\log_e q(x, t_0) + \frac{ln^3}{6} (3a - bn + 3bx)
\]

In the example given in the text the constants are as follows: \( a = 0.02 \), \( b = 0.002 \) in both area A and area B; \( k = 1 \), \( l = + \frac{1}{12} \) in A and \( k = 2 \), \( l = - \frac{1}{12} \) in B.

Annex 3

FORMULAE AND COMPUTATIONS FOR COMPARING LEVELS OF INFECTION RATES IN FIVE EUROPEAN COUNTRIES

Denoting the force of infection in the different areas \( \rho_1 (x, t) \), \( \rho_2 (x, t) \), \ldots \) it is assumed that

\[
\rho_1 (x, t) = k_1 \rho (x, t)
\]

\[
\rho_2 (x, t) = k_2 \rho (x, t)
\]

etc.

This implies that

\[
q_1 (x, t_0) = (q(x, t_0))^{k_1}
\]

\[
q_2 (x, t_0) = (q(x, t_0))^{k_2}
\]

etc.

or

\[
-\log q_1 (x, t_0) = -k_1 \log q (x, t_0)
\]

\[
-\log q_2 (x, t_0) = -k_2 \log q (x, t_0)
\]

etc.,

\( t_0 \) being the time the tuberculin-testing was done.

The average curve for Europe was obtained by computing a simple (unweighted) average of \( -\log q \) for Austria, Czechoslovakia, Greece, Poland and Yugoslavia:

\[
-\log q(x, t_0) = \frac{1}{5} \sum_{y=1}^{s} (-\log q_y (x, t_0)).
\]

Estimates of the \( k \) values were obtained by minimizing

\[
\sum_{x=1}^{18} (-\log q_y (x, t_0) + k_y \log q(x, t_0))^2,
\]

which gives

\[
k_y = \frac{\sum_{x=1}^{18} -\log q_y (x, t_0)}{\sum_{x=1}^{18} -\log q(x, t_0)}.
\]

The computations for Czechoslovakia are shown in the table below as an example.
<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Observed values for Czechoslovakia</th>
<th>Average for 5 countries</th>
<th>Computed values for Czechoslovakia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1-q$</td>
<td>$q$</td>
<td>$-\log q$</td>
</tr>
<tr>
<td></td>
<td>$\text{age correction, } -0.1$</td>
<td>(exact age)</td>
<td>$\text{-log } q$</td>
</tr>
<tr>
<td>1</td>
<td>0.038</td>
<td>0.961</td>
<td>0.017</td>
</tr>
<tr>
<td>2</td>
<td>0.051</td>
<td>0.947</td>
<td>0.024</td>
</tr>
<tr>
<td>3</td>
<td>0.070</td>
<td>0.928</td>
<td>0.032</td>
</tr>
<tr>
<td>4</td>
<td>0.092</td>
<td>0.906</td>
<td>0.043</td>
</tr>
<tr>
<td>5</td>
<td>0.112</td>
<td>0.885</td>
<td>0.053</td>
</tr>
<tr>
<td>6</td>
<td>0.140</td>
<td>0.867</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>0.171</td>
<td>0.826</td>
<td>0.083</td>
</tr>
<tr>
<td>8</td>
<td>0.200</td>
<td>0.797</td>
<td>0.099</td>
</tr>
<tr>
<td>9</td>
<td>0.229</td>
<td>0.768</td>
<td>0.115</td>
</tr>
<tr>
<td>10</td>
<td>0.259</td>
<td>0.738</td>
<td>0.132</td>
</tr>
<tr>
<td>11</td>
<td>0.285</td>
<td>0.712</td>
<td>0.148</td>
</tr>
<tr>
<td>12</td>
<td>0.317</td>
<td>0.675</td>
<td>0.171</td>
</tr>
<tr>
<td>13</td>
<td>0.369</td>
<td>0.600</td>
<td>0.222</td>
</tr>
<tr>
<td>14</td>
<td>0.437</td>
<td>0.559</td>
<td>0.253</td>
</tr>
<tr>
<td>15</td>
<td>0.475</td>
<td>0.520</td>
<td>0.284</td>
</tr>
<tr>
<td>16</td>
<td>0.524</td>
<td>0.471</td>
<td>0.327</td>
</tr>
<tr>
<td>17</td>
<td>0.577</td>
<td>0.418</td>
<td>0.379</td>
</tr>
<tr>
<td>18</td>
<td>0.627</td>
<td>0.369</td>
<td>0.433</td>
</tr>
<tr>
<td>19</td>
<td>0.688</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>...</td>
<td>...</td>
<td><strong>2.882</strong></td>
</tr>
</tbody>
</table>

Column 2 gives the proportions of positives by age: they are simple averages of those for males and females given in the report of the Czechoslovakian campaign. The age correction ($-0.1$ year) indicates that the average age for those stated as $n$ years in column 1 was $n - 0.1$. Column 3 shows the proportion of negatives at exact age $n$ obtained by linear interpolation, and column 4 gives the negative logarithms of these. The average of these last-mentioned figures and the corresponding figures for the four other European countries are shown in column 5. The $k$ value for Czechoslovakia is determined as

$$ k = \frac{2.882}{3.087} = 0.934, $$

and finally the proportion of positives are computed with this $k$ value as shown in columns 6-8.

In the map for Europe showing the percentages of positives at the age of 10 years, the percentages are given by certain groups. The limits defining these groups have been determined by setting $k = 0.2963, 0.4444, 0.6667, 1.0000, 1.5000,$ and $2.2500$ (the increase from the lower limit to the higher limit of the same group corresponds to an increase of
the infection rate of 50\% ). For \( k = 1 \) we have \( -\log q \ (1.0) = 0.150 \), and thus \( 1-q \ (1.0) = 0.2920 \); for \( k = 1.5 \) we have \( -\log q \ (1.5) = 1.5 \times 0.150 = 0.225 \), and thus \( 1-q \ (1.5) = 0.4043 \); etc.

**RÉSUMÉ**

L’épreuve à la tuberculine appliquée à une population permet d’évaluer approximativement le degré d’infection tuberculeuse de cette collectivité et par conséquent l’étendue du problème que la tuberculose y pose. Ce principe est simple. Mais une question reste entière: quels renseignements, d’ordre quantitatifs, peut-on tirer des données réunies? On s’est en effet borné jusqu’à maintenant à comparer la fréquence de l’infection tuberculeuse dans divers groupes.

L’auteur s’est proposé d’analyser les données obtenues parmi des populations soumises à une seule épreuve tuberculinique, et d’évaluer, par des courbes d’infection selon l’âge, les variations du risque d’infection durant la petite enfance, la période de scolarité, etc.

Il montre que les niveaux des taux d’infection peuvent, dans certains cas, être comparés grâce à ces courbes. Là où l’infection semble s’être développée de façon semblable, les courbes d’infection par âge permettent de comparer les niveaux d’infection par région, densité de population, groupes socio-économiques. Dans les cas où l’on ne dispose pas d’autres statistiques relatives à la tuberculose, ces courbes donnent une indication utile de l’importance relative du problème de la tuberculose. La valeur de ces courbes est cependant limitée. L’auteur suggère d’autres types de méthodes d’investigation et discutent leur valeur.

Les résultats d’épreuves à la tuberculine recueillis dans 16 pays, au cours de campagnes de vaccination au BCG de 1948 à 1951, ont servi de base à cet article. En annexe figurent le détail de l’analyse statistique et les formules appliquées pour établir le rapport entre le taux d’infection et les courbes d’infection par âge.

**REFERENCES**

1. Great Britain, Medical Research Council (1952) *Lancet*, 1, 775
2. Leopold, L. (1941) *Ned. T. Geneesk.*, 85, 1661
5. Toda, T. & Takeya, K. (1955) *Studies on purified tuberculin protein*. In: Recent investigations on purified tuberculin, Tokyo
6. WHO Tuberculosis Research Office (1954) *Mass BCG vaccination campaigns 1948-1951 in Czechoslovakia, Poland, Syria, Israel, Malta, Tunisia, Ecuador, Austria, Morocco, Tangier, Greece, Yugoslavia, Egypt, Algeria, Finland, Lebanon, Palestine Refugees, Copenhagen*
7. WHO Tuberculosis Research Office (1955) *Bull. Wld Hlth Org.*, 12, 63