value of a reality has been tested by a mathematician and confirmed......". Usually in statistics far greater probabilities are required before one is inclined to assume certainty.

In the case of the 530 male Armenians of Fig. 2 the difference between the frequencies of classes 84 and 85 is $3 \pm > 10.20$. Hence this difference is less than '29 times its standard-deviation. The odds against reliability of the peak at class 84 consequently are about 1 to $1\frac{1}{2}$. In comparing the frequencies of classes 86 and 85 the difference appears to be $10 \pm > 10.46$; the difference is less than '96 times its standard-deviation. The odds against reliability of the peak found at index-class 86 therefore are about 1 to 5.

From these probabilities one is compelled to conclude that it is far from certain that the two main tops occurring in all three Armenian cephalic index frequency-curves of fig. 2 are typical of this population. It is especially remarkable that in increasing the number of observations the odds against reliability of these peaks appear to increase too. Therefore there is a not unconsiderable possibility that the real frequency-curve of the cephalic-index in the Armenians should not show two tops and that their presence in the groups investigated must be ascribed to fluctuations of sampling, if not to accessory circumstances which favour an increase respectively a decrease of the frequencies of particular index-classes. The latter possibility will form one of the subjects of a following paper.

In later papers KAPPERS has called the peaks occurring at index-classes 83 and 86 "associated index peaks", because both of them, together or separately, characterize his Central-Asiatic race [1]). Therefore the 83 and 86 peaks should each represent a subracial group.

In frequency-curves of other populations peaks may occur which according to KAPPERS should be ascribed to different races. In case such peaks do not lie too far from each other their reliability can be investigated in the manner described here; if their distance is larger another way must be followed which will be discussed in a following paper .

[1]) Proc. Kon. Akad. v. Wetensch., Amsterdam, **37**, 612 (1934); **38**, 686 (1935).

Physics. — *Determination of the rate of infection in tuberculosis.* By Dr. G. C. E. BURGER, head of the Health Service of the N.V. Philips at Eindhoven, and Dr. H. C. BURGER, Lecturer in physics at the University of Utrecht.

(Communicated at the meeting of May 28, 1938.)

The tuberculin reaction according to PIRQUET and its later modifications (including tuberculin ointment reaction according to MORO, HAMBURGER) have not only been found to be of great significance for the diagnosis of tubercular affections, but have also made it possible to obtain with the aid of statistics an insight into the spreading of tuberculosis among the different classes of the population. The quantitative utilization of these statistics of tuberculin reaction has practically always been confined to comparing the percentage of the positive (and negative) tuberculin reactions for different classes of the population.

Recently an important attempt has been made by MUENCH in America and by STRAUB and particularly by HEIJNSIUS VAN DEN BERG in Holland to obtain from the data regarding the sensitivity to tuberculin, with the aid of mathematics, some information regarding the rate of infection by tuberculosis [1]).

If at a certain moment a group of persons of different ages is examined by means of a tuberculin test and a graph is made of the percentage of positive reactions as a function of the age, great regularity will be noticed here. This phenomenon caused HEIJNSIUS VAN DEN BERG to arrive at the conclusion that there existed a constant "factor". He assumed that this constant factor was formed by a constant rate of infection during a number of years of life. On the basis of a uninfected persons, and assuming that annually x % of these individuals becomes infected, there are then after 1 year:

$$a - \frac{x}{100}\,a \text{ or } a\left(1 - \frac{x}{100}\right),$$

after 2 years:

$$a\left(1 - \frac{x}{100}\right)^2$$

[1]) H. MUENCH, Journal of the American Statistical Association, Vol. XXIX, p. 25 (1934). — STRAUB, Report of the Genootsch. tot bev. v. Nat., Genees. en Heelk. Ned. Tijdschr. v. Geneesk., p. 1540 (1936). — HEIJNSIUS VAN DEN BERG, Report of meeting of the Ver. Ned. T. b. Artsen. N. T. v. Geneesk., p. 1872 (1937). — Reports of the Dutch Tuberculosis Committee (1937).

and after y years:

$$a\left(1-\frac{x}{100}\right)^{y}$$

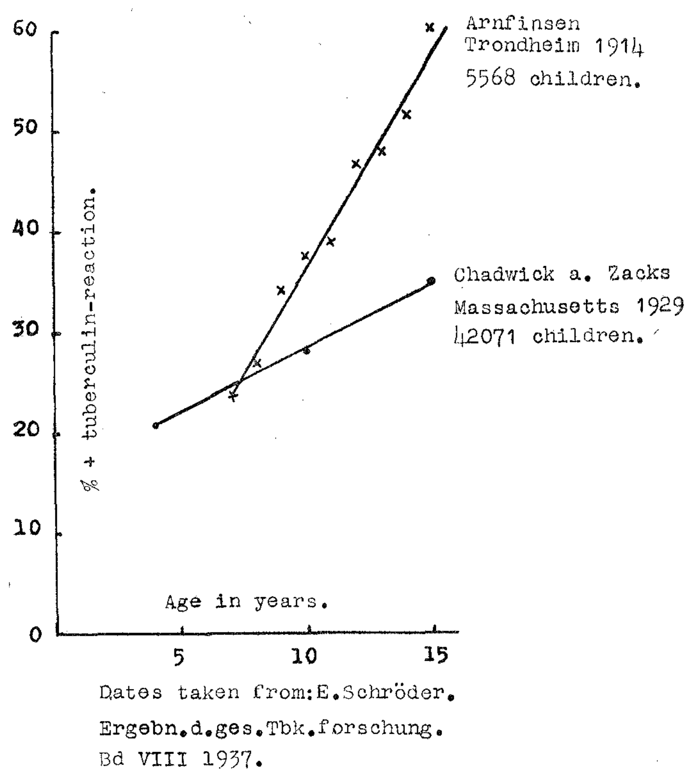
Now HEIJNSIUS VAN DEN BERG uses his formula in two different ways, viz.:

1. From a certain percentage of positive tuberculin reactions at a certain age, he calculates the annual percentage of infection, assuming that this has remained constant from birth.

2. From the comparison of the percentage of positive reactions for two age groups, he calculates the infection percentage belonging to this age group. Here again it is assumed, that the rate of infection has remained constant during the period and that the differences formed between the two ages are explained by modification of the rate of infection with the age at a certain time.

The most direct method of establishing the rate of infection would be the following. Taking as a basis a sufficiently large group of individuals

Fig. 1.

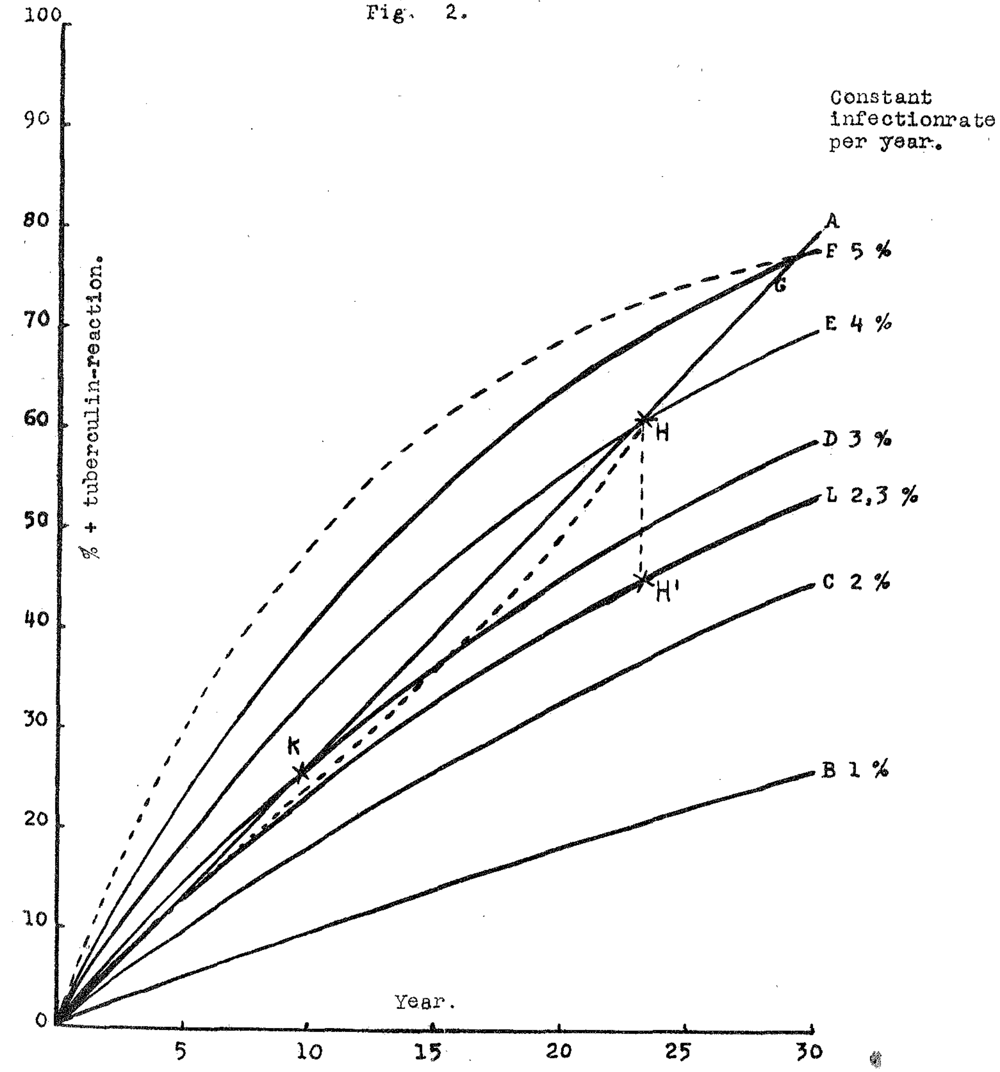


with negative tuberculin reaction, an investigation is made regarding how many persons of this group get a reversal of their reaction in the course

of a year. As far as we know, there are no data available of sufficiently large numbers of individuals with a "normal" infection rate.

It is necessary to ask oneself whether the premise regarding the existence of a constant rate of infection is really correct. The reply must certainly be negative and this is, why we think the formula from MUENCH and of HEIJNSIUS VAN DEN BERG as incorrect. When a tuberculin curve is plotted according to the age at a certain time (year) one is struck by the regular increase with age of the number of persons with positive reaction. When shown graphically on millimetre paper the line in this case is practically linear (fig. 1). If there were a constant rate of infection in the course of years an exponential curve would be the result, as in fig. 2 (continuous lines). If the real proportions of the increase in positive tuberculin reactions

Fig. 2.



according to age at a point of time is indicated by a straight line *this signifies that in the higher age groups more positive reactions occur than must be expected according to a constant rate of infection from birth.*

Fundamentally there exist two possibilities for this:

1. At the higher ages there are more infected persons because formerly the rate of infection was greater for these individuals when they were young than it is *now* for the young people; the cause therefore lies far back in the past.

2. At older ages the rate of infection is greater at a certain point of time than at younger ages; the cause is therefore to be found in more recent events.

The first possibility is shown graphically in fig. 2, where OA shows a trend, actually found, of the number of positively reacting persons according to age at a certain point of time (increase of positive reaction by 2.6 % per year), OB a constant risk of infection of 1 % per year, OC, OD, OE and OF respectively 2, 3, 4 and 5 % per year. Point G may then be imagined as resulting from an infection risk of 5 % per year from the year of birth, point H of 4 % and point K of 3 %. In reality the proportions are different because also for the age group of ± 29 years (respectively ± 23 years and 10 years) to which points G, H and K refer, the fact applies that the risk of infection was higher during the first years after birth than later. This point K can therefore be imagined as being attained by the broken line OF, which is now no longer an exponential curve.

The second possibility can likewise be seen from the figure.

With the rate of infection of 2.3 % per year according to time and age, the point H^1 at 23 years of age would be found of the line OL; in reality H is found, because with increasing age the rate of infection has increased, resulting in a larger number of infected persons, indicated for instance at an age of 23 by the distance H^1 H, whilst one can imagine this point as being reached by the dotted line OH. In HEIJNSIUS VAN DEN BERG's considerations practically only this second possibility has been taken into account. Yet the great decrease in tuberculosis infection of the population during the last decade shows clearly, that it is just this rate of infection which is diminishing, so that a method for calculating the rate of infection must certainly take this circumstance into account.

It will be clear from this consideration that it is of fundamental significance to endeavour to find a more general solution for the calculating of the rate of infection. A surveyable and certain treatment of the by-its-very-nature statistical problem is only possible by using the tried methods of statistics, mathematical formulation being unavoidable.

The considerations following hereafter apply for a large group of persons of all ages selected at random. If we ask ourselves how many of these individuals had a certain age at a given time, for instance exactly 16 years, 0 days, 0 hours, 0 minutes, 0 seconds, their number will be

found to be practically zero. Such a case is so improbable, that we need not consider it. We must therefore consider a certain range of ages. Among the 16-year olds we could for instance understand all persons between 15½ and 16½. The range of one year in the age is arbitrary. If a range of one month is allowed one would have to ask after the persons aged between 16 years — ½ month and 16 years + ½ month.

As the range is $^1/_{12}$ of those admitted first, this number will also be $^1/_{12}$ of the first number. The first number therefore entirely characterizes the "number of 16-year olds" and is indicated by n (16).

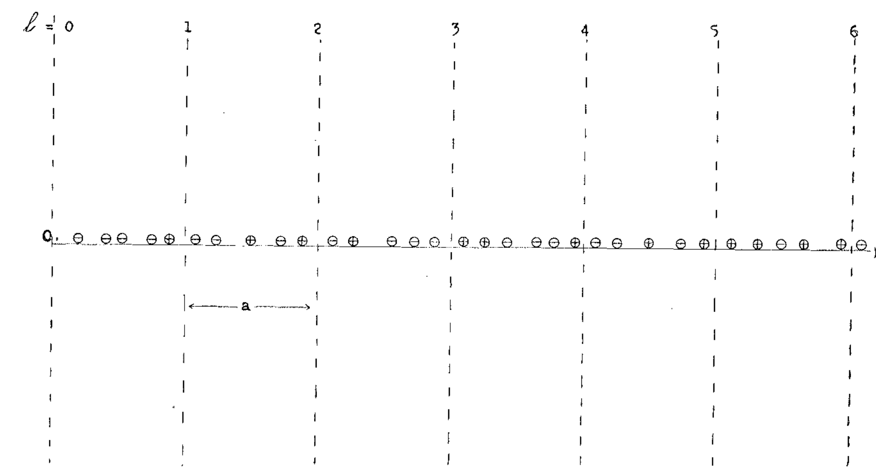
Among the $n(l)$ l-year olds there are $n_+(l)$ with a positive and $n_-(l)$ with a negative tuberculin reaction. The number of persons with positive reaction in the above fraction is therefore:

$$P(l) = n_+(l)/n(l), \text{ whilst } N(l) = n_-(l)/n(l)$$

represents the fraction with a negative reaction. It must be remembered that all the magnitudes mentioned here (n, n_+, n_-, P and N) not only depend on the *age* l, but also on the *time*, i.e. *the year*.

We must now ask ourselves what causes the magnitudes n_+ and n_- to increase or decrease in the course of time. We shall discuss these causes successively and take them into account. To make one of the most important effects clear (and it is this very effect that has been overlooked by HEIJNSIUS VAN DEN BERG) we shall use the following geometrical representation (fig. 3). At a certain time, for instance January 1st 1920,

Fig. 3.



the condition for the group of the population under consideration is shown by placing, for each individual, a point on a horizontal line at a distance from the origin 0 corresponding to the age. This point is indicated by + or —, according as the individual has a positive or negative reaction. There are just a few points in the drawing, but the statistic is only of value for many individuals, so that the actual number of points must be very great. Such a drawing gives the above-mentioned magnitudes n_+, n_- and n as the number of $+$ and $-$ signs, respectively the total number of signs on a piece of the line whose length a represents the unit of age (—difference), and therefore a year for example. The representative points ($+$ and $-$ signs) now all run with time to the right (for each individual the age increases with time), namely per year over the above-mentioned length a which represents one year. Besides this moving up with the time, the above mentioned important effect for the explanation of which we require fig. 3, our figure, which is the picture of the population group under consideration, shows much more. By birth individuals originate at $l = 0$, i.e. the origin 0 is a source of representative signs, viz., $-$ signs, which immediately start to move to the right with the uniform speed of all points. This occurrence of individuals with negative reaction at 0 is not of importance for our problem. Of primary importance, however, is that a $-$ sign may change into a $+$ sign whilst the reverse is so seldom the case that we will neglect it. Furthermore individuals, i.e. $+$ and $-$ signs in fig. 3, will disappear through decease. Finally $+$ and $-$ signs will disappear and appear through emigration and immigration.

All these effects must be taken into account and weighed up against each other in order to finally hold over what is quantitively the most important.

Through the points of fig. 3 which are generally not evenly distributed, shifting to the right, their density alters near a certain point corresponding to a certain age. If nothing else happened, the points for the age 2 ($1\frac{1}{2}$—$2\frac{1}{2}$ years) would, for instance after a year, come near the age 3 ($2\frac{1}{2}$—$3\frac{1}{2}$ years). The number of 3-year old persons with positive reaction therefore changes annually by a number corresponding to the difference between the 2- and 3-year olds and in fact increases when there are more 2-year olds then 3-year olds. If, as is usual, the annual increase of positively reacting individuals is denoted by $\partial n_+/\partial t$ and the difference in n_+ for two age groups differing one year by $\partial n_+/\partial l$ (positive when n_+ at a greater l is greater than at a smaller l) then

$$\partial n_+/\partial t = -\,\partial n_+/\partial l \quad \ldots\ldots\ldots \quad (1)$$

We indicate by the minus sign that $\partial n_+/\partial t$ is negative (i.e. that n_+ decreases with time) when n_+ is greater for older ages than for younger ages. A relation corresponding to (1) also applies for n_-.

However, the equation (1) is incomplete. At the righthand-side there is merely the alteration of n_+ by the above-mentioned effect of "moving up the year groups". Next to this, infection must be reckoned with in the first place. As a consequence of this the number n_+ increases and n_- decreases, in both cases by the same amount because each infection increases n_+ by one and n_- decreases by one. We did not take in account the possibility, that a positive tuberculin reaction can again become negative. It seems probable that this is rather seldom, and for simplifying

the problem, we only discuss the change of a negative reaction into a positive one. The number of infections occurring per year with l-aged is proportional to their number $n(l)$ and is found by multiplying this number by the infection-rate $b(t, l)$. As also remarked by HEIJNSIUS VAN DEN BERG, this depends on the age, *but also on the time t, i.e. on the year*. Without any doubt b has considerably diminished in the course of time. The equation (1) must be completed by adding to the second term the yearly number of infections $b\,n_-$. The same number must appear in the second term of the corresponding equation for n_- with the negative sign, because, as alreary observed above, infection reduces the number n_- of non-infected persons.

Furthermore we take decease into account. This gives a reduction of n_+ as well as of n_- proportional to n_+ and n_- respectively. The proportionality factor, the death rate, is not, however, exactly the same for the positively and the negatively reacting individuals and will consequently be indicated for both cases by s_+ and s_- respectively. In the second term of the equation for n_+ and n_- the terms $-\,s_+ n_+$ and $-\,s_- n_-$ will have to be added, which represent the number of persons deceased per year.

Both s_+ and n_- depend in their turn on the age l and the time (year) t.

Finally we must also mention the effect of immigration and emigration. This althers the number n_+ per annum by an amount $\triangle_+$, the excess of settlement over departure, whilst n_- varies with $\triangle_-$. Not much more need be said about these amounts, because they are greatly dependent on local circumstances.

Resuming, therefore, the equations for n_+ and n_- read as follows:

$$\frac{\partial n_+}{\partial t} = \frac{\partial n_+}{\partial l} + b\,n_- - s_+ n_+ + \triangle_+ \quad \ldots\ldots \quad (2)$$

$$\frac{\partial n_-}{\partial t} = \frac{\partial n_-}{\partial l} - b\,n_- - s_- n_- + \triangle_- \quad \ldots\ldots \quad (3)$$

Increase of number of positively or negatively reacting individuals with time	moving up of the year-groups	infection	mortality	removal

For comparison with the considerations of HEIJNSIUS VAN DEN BERG it must be observed here that he quite rightly leaves mortality and removal out of consideration. But he furthermore assumes that the state of affairs is stationary, i.e. that it does not change with time. This amounts to setting at zero the left-hand side of this equation [1]). It will be seen later that this is not allowable.

It is necessary to deduce from equations (2) and (3) applying for the

[1]) The equation (3) can therefore be solved in a simple manner and it is this solution that is used by HEIJNSIUS VAN DEN BERG.

numbers n_+ and n_- of the persons with respectively positive and negative reaction, a relation for the fraction already mentioned $P = \frac{n_+}{n}$ of all persons of the age l having a positive reaction. As a matter of fact this fraction is the magnitude that observation gives or should give us. The calculation that leads to the relation P is of no importance for understanding the problems dealt with.

The result is that we find the required rate of infection b expressed as follows in observable magnitudes:

$$\underset{\text{rate of infection}}{b} = \underset{\text{increase with age}}{\frac{1}{1-P}\frac{\partial P}{\partial l}} + \underset{\text{increase with time}}{\frac{1}{1-P}\frac{\partial P}{\partial t}} + \underset{\text{mortality}}{P(s_+ - s_-)} + \underset{\text{removal}}{P\left(\frac{\Delta_-}{n_-} - \frac{\Delta_+}{n_+}\right)}. \quad (4)$$

We shall discard the last term. As a rule nothing can be said of the "removal term" and there is therefore no reason to assume that those who have removed differ in composition from those present.

The term referring to mortality is certainly positive, because the mortality-rate of the positively reacting persons s_+ is greater than that of the negatively reacting individuals s_-.

As, however, s is of the order of magnitude of a few per thousand per annum and a small fraction of the mortality is due to tuberculosis, we find for the "morality term" in (4) a value of only a few tenths per thousand. As the other terms are of the order of a few per cent per annum, the about 100 times smaller "mortality term" can be discarded with respect to them.

Finally the equation (4) condenses to the following simple relation for the rate of infection b:

$$b = \frac{1}{1-P}\frac{\partial P}{\partial l} + \frac{1}{1-P}\frac{\partial P}{\partial t} \text{ or: } \boxed{b = \frac{1}{1-P}\left(\frac{\partial P}{\partial l} + \frac{\partial P}{\partial t}\right)} \quad . \quad (5)$$

If P, the fraction of those reacting positively, were known as a function of the age l and the year t, the problem would be solved in this way. But, unfortunately, the observation material is so scarce and at the same time so contradictory that we shall have to confine ourselves to a rough estimation of b.

Unfortunately, we shall have to consider as unattainable for the present the ideal case, namely establishment of b as a function of the age and year, for different countries and groups of population.

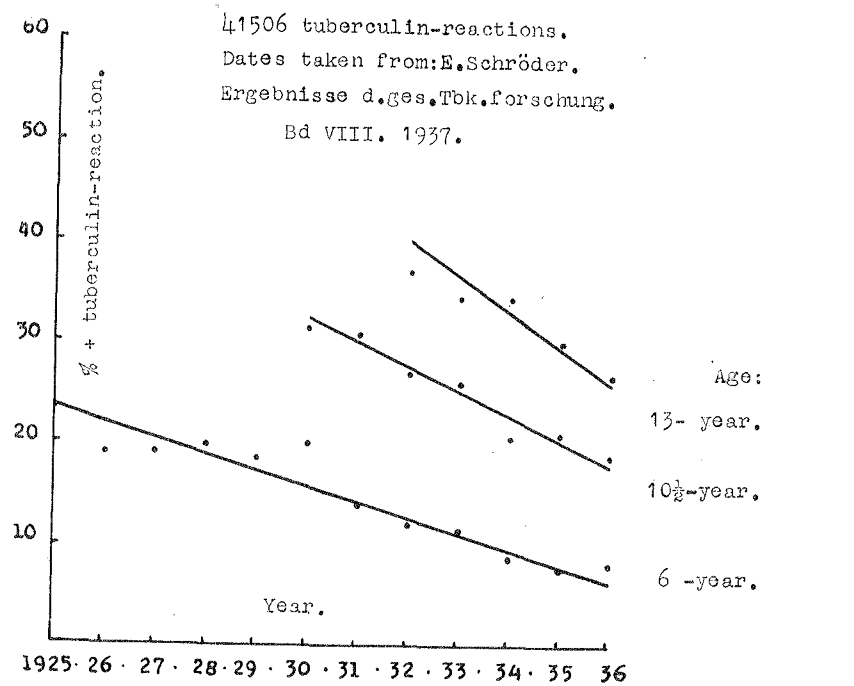
In fig. 4 we give a graphic representation of the percentage of children with positive reaction as a function of the year, according to E. SCHRÖDER. The slope of the line is greater at an older age, but in connection with the uncertainty of the data we shall confine ourselves for the present to the average life of 10 years and a time corresponding to the year 1930.

The slope then gives us for the reduction of P with the time:

$$\frac{\partial P}{\partial t} = -2.5\ \% \text{ per year.}$$

The observations of KLEIN as well as those of PERETTI (fig. 5) give values of 2—3 % per annum for the diminution of P with the time,

Fig. 4.



although it must be remarked that, according to PERETTI, for 6-year olds the reduction is practically nil after 1932.

From figs. 4 and 5 can also be deduced the value of $\frac{\partial P}{\partial l}$ i.e. the dependence of P on age.

Thus, for instance, in fig. 4 the line for $l = 13$ is about 10 % higher than that for $l = 10\frac{1}{2}$, from which there follows for $\frac{\partial P}{\partial l}$ a value of $\frac{10}{(13-10\frac{1}{2})} = 4$.

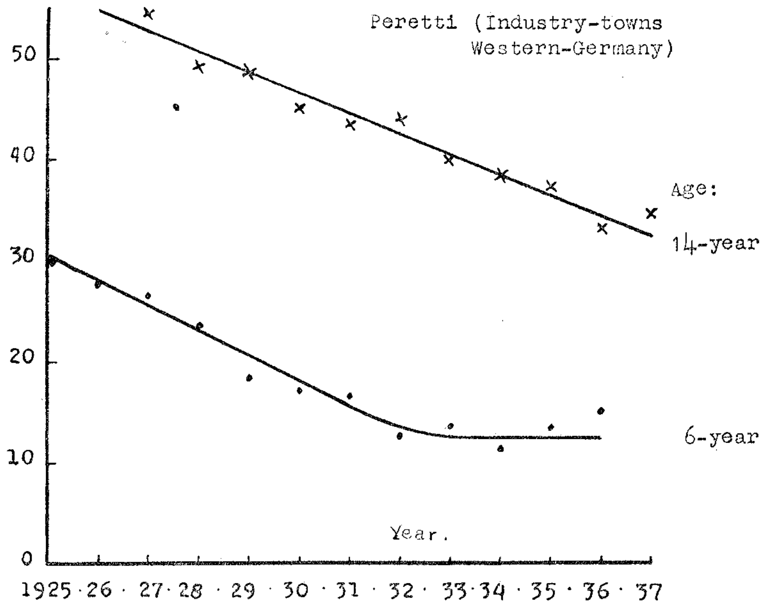
If this method of calculation is applied to the lines of figs. 4 and 5, we get as a mean value:

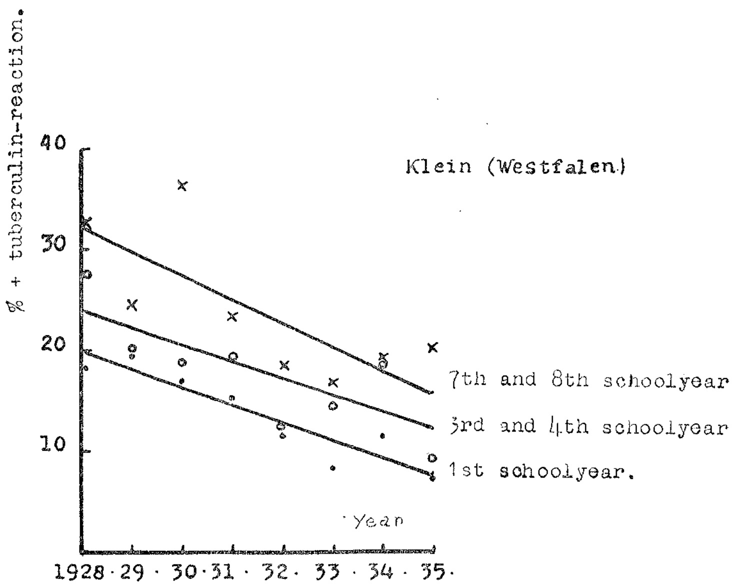
$$\frac{\partial P}{\partial l} = 3.5\ \% \text{ per year,}$$

a value which again applies for the approximately 10-year olds and for the year 1930.

It must not be forgotten here that the values of $\frac{\partial P}{\partial l}$, found from the different pairs of lines, differ considerably (from about 1.8 % to about

Fig. 5.



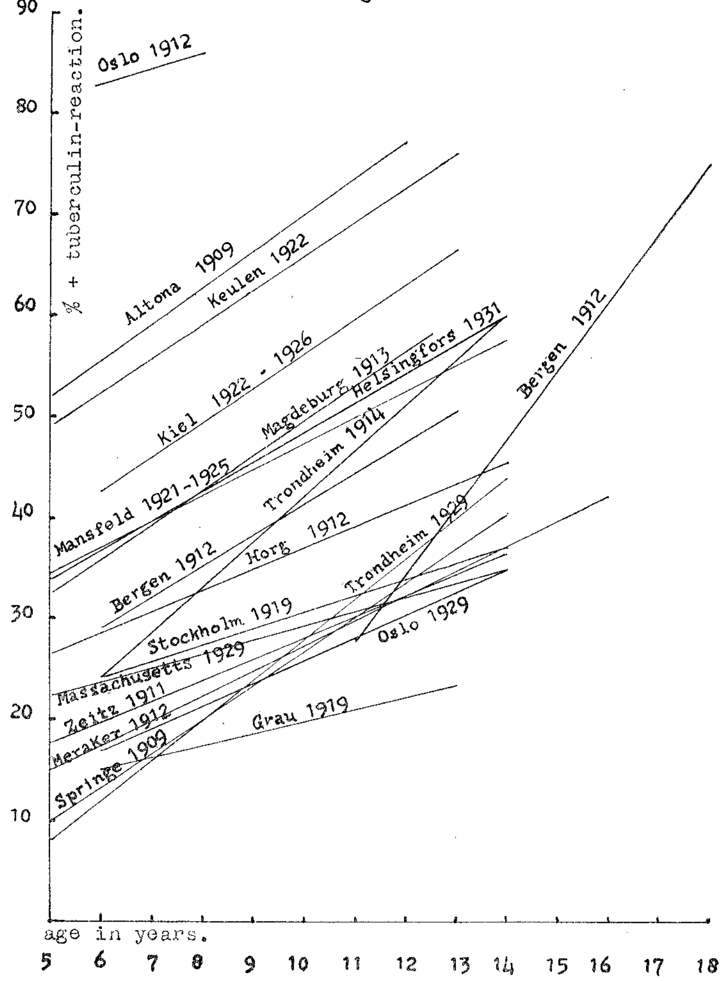


Dates taken from E.Schröder,
Ergebn.d.ges.Tbk-forschung.Bd VIII.

4.4 %), so that the above value of $\left(\frac{\partial P}{\partial l}\right)\begin{matrix} l = 10 \\ t = 1930 \end{matrix}$ must merely be considered as a rough estimation.

The great uncertainty appears very clearly from fig. 6, in which P is now represented as a function of the age l.

Fig. 6.



Dates taken from Mattison. Ergebn.d.ges.Tbk.forschung.
Bd. V.

Here, with one exception, each line is established by only two points and therefore we cannot do otherwise as to draw a straight line between these points.

By their slope the lines give $\frac{\partial P}{\partial l}$ and the values for $\frac{\partial P}{\partial l}$ vary between 1.5 to 6.5 % per year. This does not conflict with the above-mentioned value of 3.5 % per annum, but is of little value next to the more reliable data.

Measurements of the same group of the population in different years only occur once here (TRONDHJEIM). This gives for $\frac{\partial P}{\partial l}$ a value of 1 % per year. However, not too much value must be attached to the latter, because it is only one of the many lines.

The rate of infection b for 10-year olds in 1930 for a number of large towns in Germany finally follows from equation 5, applied to the graphs of figs. 4 and 5:

$$b = \frac{1}{1-0.30}\,(3.5-2.5) = 1.5\ \%\ \text{per year.}$$

From this it can be seen that neglection of the term $\frac{\partial P}{\partial t}$, the moving up of the year groups, would give an entirely incorrect result of

$$b = \frac{1}{1-0.30} \times 3.5 = 5\ \%\ \text{per year.}$$

Actually, therefore, the consideration of MUENCH and of HEIJNSIUS VAN DEN BERG, who only take into account the dependence of P on the age l, is not only fundamentally, but also quantitatively, incorrect, as was already expected at the beginning of this article on the basis of more qualitative considerations.

There are of course many cases where the calculation of HEIJNSIUS VAN DEN BERG gives a good approximation. If the figures for TRONDHJEIM may be relied upon, it then follows from them that $\frac{\partial P}{\partial l} = 4.5$ % per year and $\frac{\partial P}{\partial l} = 1$ % per year, therefore $b = \frac{1}{1-0.3}\,(4.5-1) = 5$ % per year (for 10-year olds in the period 1914—1929).

Here omission of the term $\frac{\partial P}{\partial l}$, would give a fairly good result of $\frac{1}{1-0.30} \times 4.5 = 6.5$ % per year.

It is also possible to endeavour to arrive at an estimation of the rate of infection for other ages:

$$b_{l=12} = \frac{1}{1-0.35}\,(4-2.8) = 2\ \%\ \text{per year,}$$

$$b_{l=8} = \frac{1}{1-0.25}\,(3.5-2.2) = 1.7\ \%\ \text{per year}^{1)}.$$

It is doubtful whether the somewhat higher value that is found for the rate of infection at older ages has a real significance. It is also certain that discarding $\frac{\partial P}{\partial t}$ would give an entirely incorrect result, as follows from

1) These figures have been deduced from SCHRÖDER's observations only (fig. 4), which is why the values are slightly larger than those for 10-year olds mentioned above.

the calculation by HEIJNSIUS VAN DEN BERG, who found large differences in the rate of infection between these age groups.

Finally, as regards the variation of the infection rate with the time (year), it must be observed that the lines of fig. 4 come closer together towards the right. This means a diminution of $\frac{\partial P}{\partial l}$ with the time.

Assuming the lines to be straight, this would signify that $\frac{\partial P}{\partial t}$ is independent of the time. The result is consequently a reduction of $\frac{\partial P}{\partial l} + \frac{\partial P}{\partial t}$.

As, moreover, the risk of infection must be divided by $1-P$, where P diminishes with the time, there follows from both facts a decrease of the rate of infection b with the time.

Finally, a further remark of a general nature, regarding the statistical problem dealt with here. Two entirely different questions can be asked. In the first place the important magnitudes (here the rate of infection: b can be deduced from the experimental data and equation (4) which are related to the mechanism of the phenomena.

In the second place an endeavour can be made, with the aid of the equations, to give a reply to the question as to what the future will bring us. This second method of dealing with the problem, which is a very usual one in physics, astronomy, chemistry, etc. encounters great difficulties when applied to "human" problems. One would have to forecast how in the future the rate of infection b would depend on time and age in order to subsequently "solve" equation (4). While this solution is of a purely mathematical nature and in principle simple and practicable in actual practice, the forecast is a medical problem of a less simple nature. Though it may certainly be said that b depends on the number of sources of infection and will be proportional to the same, the proportionality-factor depends on many social and hygienic factors (isolation of patients, succeptibility, etc.) and although we might dare to say something about them as regards the present time, an extrapolation for the future would be hazardous.

For this reason it would seem to us that only the evaluation from the available data of the rate of infection in the past and in the present is of any value. To make it possible in future to have correct data available for treating tuberculin statistics in the right way, it is necessary to establish the percentage of positive reactions for every calendar year and for every age group.

Summary.

From the statistics of the tuberculin reaction a closer insight can be obtained into the greatness of the rate of infection according to age and year. For the evaluation of this rate of infection a simple relation is deduced.