Annual Risk of Tuberculosis Infection in the Transkei

PROBLEMS CONNECTED WITH ITS ESTIMATION FROM THE DATA OF A TUBERCULIN SURVEY

C. DE VILLE DE GOYET

SUMMARY

The tuberculin and X-ray survey undertaken in the Transkei by the State Health Service, SANTA, and the Tuberculosis Research Unit of the South African Medical Research Council was used for a study of the prevalence of tuberculous infection in children from 3 months to 15 years old. The tuberculin test considered was the Mantoux test with 2 TU of human PPD, and an induration of 10 mm was regarded as indicative of a previous tuberculous infection. The combined trend of the risk with age and calendar years was estimated by a mathematical method. The annual risk of tuberculous infection was calculated to be approximately 7%. A downward trend of only 2% per year was seen.

S. Afr. Med. J., **48**, 957 (1974).

A tuberculin survey performed on a representative sample of the population, i.e. a randomly selected sample, provides a good amount of information on prevalence of tuberculous infection. As pointed out by Nyboe,[1] 'a disadvantage of the information provided by a simple tuberculin survey at different ages is that the analysis does not permit a separation of the calendar trend in the risks of tuberculous infection from any relationship between age and the risk of infection. The distinctive feature of such a survey is that it gives the prevalence of tuberculin positivity in children and adolescents over a wide age range at the same point of time'.

In the Netherlands the calendar trend in the risk was found by the Tuberculosis Surveillance Research Unit[2] (TSRU) of the International Union Against Tuberculosis (IUAT) to have been of considerably greater importance than the association between age and the risk, so Sutherland *et al.*[3] have theoretically disregarded any association of the risk of infection with age in their analysis of the surveys performed in 14 countries of Europe and North Africa. From further studies, so far unpublished, these authors considered that it is not justifiable to ignore the possibility of an association between the risk of infection and age in the majority of countries.

Tuberculosis Research Unit of the South African Medical Research Council, Onderstepoort, Tvl
C. DE VILLE DE GOYET, M.D.

Date received: 26 November 1973.

TUBERCULIN TESTING

A random survey of the population of the Transkei was undertaken by the State Health Services with the co-operation of the Transkei Branch of SANTA, and the Tuberculosis Research Unit of the South African Medical Research Council. In all, 1 458 children, aged 3 months to 15 years, were tuberculin-tested simultaneously by the multiple puncture method and a double Mantoux test using human and avian purified protein derivative (PPD). A detailed report was published.[4]

The reactions to a low dose of 2 TU human intradermal PPD will be considered in this study.

A mean induration (an average of the readings by 2 independent observers) equal to, or more than, 10 mm was considered as a positive result, indicating past tuberculous infection with human tubercle bacilli.[4] The prevalence of human tuberculous infection might be overestimated by the criterion because of the frequency of non-specific myobacterial infection in some of the sites surveyed.

METHODS OF ANALYSIS

Several methods of estimating an apparently simple parameter, namely the incidence of tuberculous infection, expressed as an annual risk of infection, from the data of a single vertical survey, have been published in the last 10 years. Five of them were applied with varying success to the data from our survey, and for the sake of clarity, they will be referred to as methods 1 - 5, respectively.

Method 1

An easy, but inaccurate, estimate may be obtained by dividing the prevalence rate of positive reactors by the age of the children on whom a Mantoux test was done. This method does not take into account the usual feature of the age-infection curve, which shows a steady rise with age, levelling off or even slightly declining after attaining a maximum. Indeed, we implicitly assume a linear progression of the prevalence rate with age. This shortcoming might be acceptable in a survey of young children in a

low prevalence area.[5] Furthermore, no separate estimate of the annual risk of infection for each calendar year may be obtained. For instance, the prevalence of infection in 6-year-old children divided by the age will give us an estimate of the annual risk of infection over a period of 6 years. For the aforementioned reasons, the results obtained by this method were not considered in this study.

Method 2

This method is based on the following: if there was no association between the risk of infection and the age of the child, then the difference in experience of tuberculous infection between, say, the group aged 2 years at the time of the survey and the group aged 1 year, is the experience of infection in the latter during the calendar year preceding the birth of the younger cohort. In our survey, conducted in January 1972, the cohort aged 2 years (average $2\frac{1}{2}$ years, since this group is supposed to include children aged exactly 2 years to just under 3 years, and for convenience will be considered as aged exactly 2 years) is born in January 1970 and the cohort aged 1 year (average) is born in January 1971. A comparison of the prevalence rates of positive reactors at the ages will therefore provide an estimate of the risk of infection in 1970.

An algebraic expression of the relationship between prevalence rate at different ages is better expressed in terms of prevalence of non-infected children or non-reactors.

The probability of not being infected at any given age is the product of the probability of not being infected at every year since birth date. Let us take a simple example and consider the annual incidence of infection to be constant, say 5%; the probability of being infected every year is therefore 0,05 and that of remaining uninfected is obviously 0,95. In a cohort of 1 000 children aged 2 years the probability of being uninfected at birth (birth date 1 January 1970) was 1,0. After 1 year (January 1971) 0,95 × 1 000 will remain uninfected, that is 950 out of the 1 000. After a second year of life (January 1972, time of the survey), assuming that there is no reversion to negativity, out of the group of 950 uninfected children the only one of interest 0,95 × 950 = 0,95 × 0,95 × 1 000 = 902,5 (i.e. prevalence of infected = 8,75%) will remain uninfected at the time of the survey. The algebraic expression is therefore $Q_{1972,\ 2\ years} = Q_{1971} \times Q_{1970}$.

A numeral and hypothetical example will elucidate this rather abstract concept. Let the prevalence of reactors in a cohort aged exactly 2 years tested in 1972 be $P_{1972} = 0{,}40$ (i.e. 40% of positive reactors) compared with 0,20 in the cohort aged 1 year. What is the risk of infection in 1970, assuming no association between age and risk of infection?

$Q_{1972,2} = 1 - P_{1972,2} = 0{,}6$ (prevalence of non-infected children aged 2 years in 1972)

$Q_{1972,1} = 1 - P_{1972,1} = 0{,}8$ (prevalence of non-infected children aged 1 year in 1972)

$\frac{Q_{1972,2}}{Q_{1972,1}}$ = 'risk' of remaining uninfected in 1970 = $Q_{1970} = \frac{0{,}6}{0{,}8} = 0{,}75$.

The risk of tuberculous infection in 1970 is therefore estimated to be 1 − 0,75 = 0,25. The risk of infection in other years may be estimated similarly. Unless very large numbers of children are tested, the estimate obtained may be very imprecise.[3] In estimating risk of infection from surveys using age groups 0 - 4 and 5 - 9 years, this method will give an estimate of the product of the risk of infection over 4 consecutive years. By taking the geometric mean of this product, that is the 4th root using logarithmic tables, an annual estimate will be derived.

Method 3

This method[6] is based on the assumption that the change brought about in the proportion uninfected at any age corresponds closely to an exponential increase with time. No assumption is made on the trend of the annual risk itself. In other words, it follows the mathematical relationship $Q' = Qe^{-rn}$, where Q and Q′ are the prevalence of uninfected in 2 consecutive age groups, r is the annual risk of infection and n is the interval between the mid-points of the 2 age groups. In particular, for the new-born the annual risk of infection is derived from the equation where n is equal to 7,5 months, or $\frac{7{,}5}{12}$ our observed prevalence of uninfected was 0,955, hence $0{,}955 = e^{-7{,}5/\frac{12}{12}\,r}$ which gives a value of 0,074 (7,4%) for r the annual risk of infection in 1971. For persons in the age group 11 - 14 years a similar value (6,75%) for r is obtained from the equation $0{,}39 = 0{,}511\ e^{-4r}$ where 0,39 is the prevalence of uninfected in this age group (Table I), 0,511 the prevalence of uninfected in the 7 - 10 years age group and n = 4 years.

Method 4

The TSRU used a more comprehensive method to analyse the trend of tuberculous infection in 14 countries. The basic assumptions underlying the mathematical relationship are similar to those of method 3, with the addition that the annual risk of infection is expected to show a downward or possibly upward trend corresponding closely to an exponential curve. In a first stage an estimate of the constant rate of decrease of the risk is obtained by an iterative procedure (trial - error). Access to a computer is a prerequisite because of the amount of calculation involved. By comparing any pair of prevalence figures an estimate of the exponential rate of decrease or increase with time of the annual risk of infection may be derived. The best combination was selected by iteratively minimising the squared difference between observed and calculated values of prevalence. The detailed computational

techniques are thoroughly described in the original article.[3] This method is to be considered as the best, because of its basic assumption of a dynamic trend in the annual risk of infection in association with calendar time and age. These two factors may not be dissociated in analysing the results of a simple tuberculin survey.

Method 5

The same authors applied the 'least-square' method by calculating the best-fit line to the observed values of $\ln(-\ln Q)$. This method, applied to our figures, did not produce satisfactory results. The preceding method 4 is mathematically slightly simpler than this weighted least-square method.

These 5 methods were applied onto the observed figures of prevalence of non-infection, as well as onto their expected values obtained by fitting a parabola to the data by the least-square method. The geometric mean was used to determine average age and hence birth date of each age group.

RESULTS AND DISCUSSION

The prevalence rates of uninfected children per age group with the corresponding average age (geometric mean) and average birth date are shown in Table I. As the observed prevalence (Q) shows an unwanted random variability due to the rather small sample, the figures obtained from the best-fit parabola ($\hat{Q}$) were calculated and used jointly in this study.

No mention will be made of the results obtained by method 1, since the assumptions are unrealistically simple. We did not succeed in obtaining satisfactory results by using method 5.

In Table II we find the annual risk of infection as estimated by methods 2, 3 and 4, applied to the observed prevalence rate of infection. The annual risk calculated from the expected values of the prevalence rates as obtained by the best-fit parabola is similar and therefore not tabulated. Clearly, no pattern is readily visible in the first 2 columns (methods 2 and 3). Method 4 takes advantage of the full set of observations as opposed to methods 2 and 3, which compared 2 successive years. The average annual risk from observed values for the period covered by the survey (1957 - 1971) is given for methods 2 and 3, i.e. 7,2% (method 2) and 7,16% (method 3). As the methods did not assume any specific pattern to the annual risk of infection with the calendar time, we may accept the average as a valid unique estimate. The very feature of method 4 is to assume an exponential decrease of this annual risk over the time, so no over-all average of the results is mentioned.

TABLE II. RISK OF TUBERCULOUS INFECTION IN THE TRANSKEI, ACCORDING TO ASSOCIATION OF AGE AND CALENDAR YEAR

Calendar year	Age	Risk of infection (%) from observed figures: Method 2	Method 3	Method 4
1971	3 - 12 months	10,5	7,4	6,25
1970	1 year	0,0	0,6	6,36
1969	2 years	11,6	12,3	6,49
1968	3 years	9,8	10,3	6,62
1967	4 years	0,0	0,0	6,75
1966	5 years	19,5	19,6	6,88
1965	6 years	0,0	0,0	7,02
1961 – 1964	7 - 10 years	7,2	7,5	7,34
1957 – 1960	11 - 14 years	6,5	6,8	7,93
1957 – 1971		7,2	7,16	—

It may be estimated that the incidence rate of infection in the Transkei was approximately 7% during the period 1957 - 1971, i.e. every year approximately 7 out of 100 uninfected persons will be infected by *M. tuberculosis*. This is a very high rate compared with Asian and other African countries. It may be slightly overestimated because of the non-specific tuberculin sensitivity caused by mycobacteria other than tubercle bacilli occurring in the environment, which were also found in mouth cavities of apparently healthy individuals.

The more interesting part of the work is to calculate by means of method 4 any trend downward or upward in that annual risk. A downward trend means a favourable

TABLE I. PREVALENCE OF UNINFECTED CHILDREN PER AGE

Age group	Mean age (geometric mean)	Year of birth (average)	No. of children	Prevalence of uninfected: Observed Q	Best-fit parabola $\hat{Q}$
3 - 12 months	0,5	1971	69	0,955	0,979
1 year	1,4	1970	86	0,960	0,918
2 years	2,45	1969	100	0,849	0,853
3 years	3,5	1968	117	0,766	0,793
4 years	4,5	1967	110	0,767	0,736
5 years	5,5	1966	111	0,618	0,682
6 years	6,5	1965	109	0,690	0,631
7 - 10 years	8,8	1963	401	0,511	0,528
11 - 14 years	12,8	1959	352	0,39	0,387

TABLE III. OBSERVED AND CALCULATED PERCENTAGES OF TUBERCULOUS-INFECTED CHILDREN (REACTION 10 mm OR MORE THAN 10 mm TO 2 TU PPD)

		Calculate % (from observed figures)		
Age	Observed %	Method 2 (average risk 7,20)	Method 3 (average risk 7,16)	Method 4 (decrease 2% of the risk)
3 - 12 months	4,5	3,7	3,5	3,2
1 year	4,0	10,0	9,6	8,8
2 years	15,1	16,8	16,1	15,1
3 years	23,4	22,8	22,0	20,6
4 years	23,3	28,4	27,4	26,1
5 years	38,2	33,6	32,5	31,2
6 years	31,0	38,4	37,1	36,0
7 - 10 years	48,9	48,1	46,6	46,3
11 - 14 years	61,0	61,6	60,0	61,5
Sum of squared differences		142,7	128,9	108,4

impact of control measures on the tuberculosis problem. An upward or stationary trend would show the failure of previous efforts. As explained under Methods, the best estimate is expected to fit closer to the observed prevalence figures than all others. The results of method 4 were shown to be identical when applied to the observed figures or to the best-fit parabola figures which are not tabulated (Figs 1 and 2). The annual percentage of decrease in the infection risk was estimated to be 2% or less. In the TSUR study[3] of 14 countries the decrease between 1938 and 1948 was over 5% in most of the countries surveyed, and up to 13% in some.

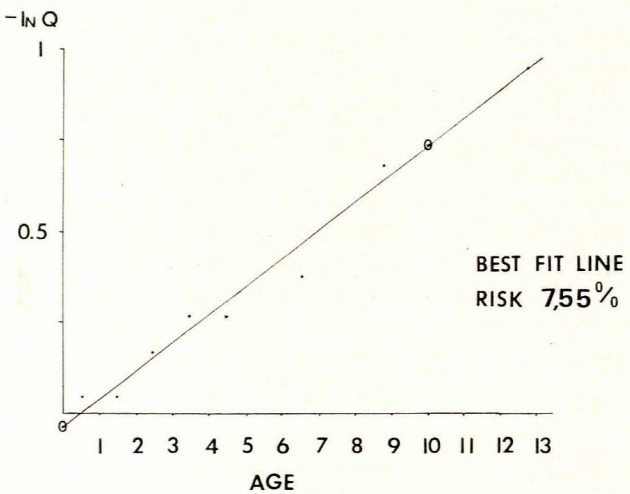

Fig. 1. Results of method 4 applied to observed figures.

In Table III the calculated percentages of positive reactors are compared with the observed figures. The sum of the squared differences is shown in the last row as a measure of fitness. The last 2 columns show a slight difference in favour of the hypothesis of an annual decrease of 2% of the infection risk. The sum of squared differences following methods 2, 3 and 4 were 142,5, 128,9 and 108,4 respectively. The method of least square (method 5), applied to the logarithm of the risk, yielded a sum of 776,5. One can accept a decrease of 2% in the annual risk as an estimate of the actual trend in the Transkei. During the various computer runs a decrease of 1%, 1,5% or 2% yielded very close results. It is not possible to sort out the separate effects of age and calendar years. Recent unpublished work of the British MRC and the TSRU has made it clear that one cannot ignore the possibility of age effect. The downward trend of about 2% a year in the risk of tuberculous infection should be confirmed by a repeat survey in the same areas after an interval of at least 5 years.

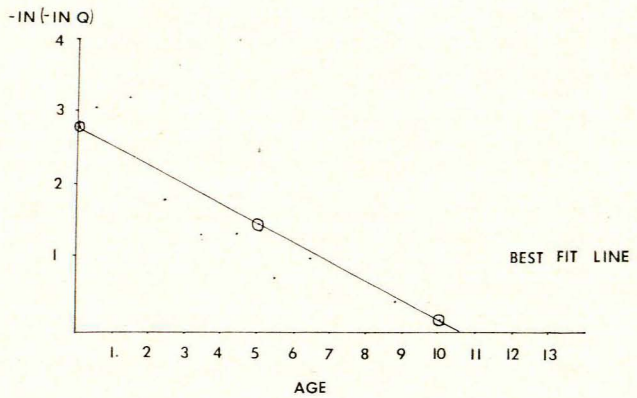

Fig. 2. Results of method 4 applied to best-fit parabola figures.

If there should be any trend of infection risk with age, it is likely to be an increase with age, i.e. the percentage decrease in risk of infection would be over-evaluated, i.e. the decrease, if any left, could actually be lower.

REFERENCES

1. Nyboe, J. (1957): Bull. Wld Hlth Org., **17**, 319.
2. Tuberculosis Surveillance Research Unit (1969): Bull. Int. Un. Tuberc., **42**, report No. 1.
3. Tuberculosis Surveillance Research Unit (1971): *Ibid.*, **45**, report No. 2.
4. SA Tuberculosis Study Group (1974): S. Afr. Med. J., **48**, 149.
5. Lotte, A., Perdrizet, S. and Hatton, F. (1971): Bull. Wld Hlth Org., **44**, suppl. pp. 1 - 228.
6. Raj Nairain, Nair, S. S., Chandrase Khar, P. and Ramanatha Roa, G. (1966): *Ibid.*, **34**, 605.