Epidemiol. Infect. (1992), **108**, 367–375
Printed in Great Britain

Methods for estimating the incidence of primary infection in pregnancy: a reappraisal of toxoplasmosis and cytomegalovirus data

A. E. ADES

Epidemiology and Biostatistics Unit, Division of Public Health, Institute of Child Health, 30 Guilford Street, London WC1N 1EH

(*Accepted 4 October 1991*)

SUMMARY

Accurate incidence information is required to plan and evaluate screening programmes which have been proposed for the detection of primary toxoplasmosis and cytomegalovirus infection in pregnancy. Appropriate statistical methods are described for deriving incidence rates and their confidence intervals from three types of data: change in age-specific seroprevalence, seroconversion, and IgM studies. These methods are applied to seven published studies on toxoplasmosis and cytomegalovirus carried out in the UK. In these publications only one estimate of the infection rate per pregnancy was correctly derived, and none were accompanied by confidence intervals. Using the proposed methods, most estimates of the primary toxoplasmosis rate in these studies were between 2·5 and 5·5 per 1000 pregnancies, compared to the 2 per 1000 usually cited. Most cytomegalovirus incidence estimates were between 4 and 10 per 1000 pregnancies.

INTRODUCTION

There has been a continuing debate on the potential of prenatal screening programmes to prevent congenital toxoplasmosis and cytomegalovirus (CMV) infection [1–8]. All congenital toxoplasmosis infection and, in the UK, most congenital CMV infection [9], follows a primary, as opposed to recurrent, maternal infection in pregnancy. Any screening programme must therefore take into account the number of women of childbearing age who are susceptible, the incidence of primary infection in susceptible women during pregnancy, the probability of transplacental transmission, and the probability that fetal infection will result in damage evident at birth or in later life. As in any screening programme, the sensitivity, specificity and positive predictive value of the tests must be taken into account, as well as the consequences of false positive and negative screening results, and the efficacy of the intervention [10].

One particularly critical link in the chain of reasoning is the incidence rate of these infections in women of childbearing age. Incidence estimates can be derived from several types of epidemiological study, but many of the published estimates have been calculated incorrectly. Methods appropriate for each epidemiological design are described, and applied to published data from UK studies.

GENERAL CONCEPTS AND METHODS

The approach recommended here is first to derive an *annual incidence*, defined as the number of primary infections per 1000 susceptibles per year. This enables incidence estimates from all studies to be compared, irrespective of the epidemiological methods used. Next, *incidence per pregnancy* should be derived, which is the number of pregnancies, per 1000, in which a primary infection occurs. Unlike the annual incidence, incidence per pregnancy includes both seropositive and seronegative individuals in the denominator. It therefore gives an indication of the proportion of pregnancies at risk, and hence the potential impact of a prenatal screening programme. Other quantities have been calculated in the literature, and these are also discussed.

ANNUAL INCIDENCE OF PRIMARY INFECTION

Changes in age-specific seroprevalence (CSP). Seroprevalence surveys on antenatal sera, or on sera collected in other ways, form the basis for this method. The rate at which seroprevalence rises with age in a cross-sectional survey can be used to estimate incidence. Such calculations make a number of assumptions [11]. Firstly, seropositivity must not decline with time after the infection. (Some possible counter-examples have been discussed for rubella [12] and toxoplasmosis [13].) Secondly, there should be no trends in seroprevalence over time. Thirdly, it is necessary to assume that older women in the survey have had the same age-specific infection rate as younger women. If the data are based on antenatal sera this may not be the case, as young women having babies tend to be demographically different from older ones. Finally, it is relevant to consider whether the annual incidence estimated from CSP, and based therefore on infection experience both during and outside periods of pregnancy, would be expected to apply to pregnant women. This may not be true of sexually transmitted infections such as CMV. It is not always possible to check any, let alone all, of these assumptions in a single cross-sectional survey, and as a result incidence estimates derived from seroprevalence data may be seriously biased.

Age-specific seroprevalence data can be modelled in terms of underlying incidence by assuming that infection acts on a declining proportion of susceptibles [14–16]. $q(a)$, the proportion of seronegatives, can be found by integration of $I(a)$, the incidence at age a:

$$q(a) = \exp\left(-\int_0^a I(s)\,ds\right). \tag{1}$$

It is reasonable to assume constant incidence I_{CSP} over the childbearing period, starting at age A, but not necessarily constant since birth. Rewriting (1):

$$q(a) = \exp\left(-\int_0^A I(s)\,ds - \int_A^a I_{\mathrm{CSP}}\,ds\right). \tag{2}$$

The first term in the exponent integrates to a constant, and

$$\ln\left(q(a)\right) = I_{\mathrm{CSP}}\,a + k, \quad \text{for } a > A. \tag{3}$$

Thus, the log proportion seronegative is linear in age over the childbearing period,

https://doi.org/10.1017/S0950268800049839 Published online by Cambridge University Press

with slope $-I_{CSP}$. If the data is grouped by age, and a_i is the mean age, with q_i the proportion seronegative in the ith group, $i = 1, 2, \ldots, n$,

$$\ln(q_i) = -I_{CSP}\, a_i + k. \tag{4}$$

Binomial regression with a logarithmic linking function [17] may now be used to estimate I_{CSP} [18]. An approximate alternative, in the absence of suitable software, would be to estimate the slope from any two points (a_1, q_1) and (a_2, q_2) on the scatterplot (4). Consider two age groups of sizes N_1 and N_2, and mean age difference $D = a_2 - a_1$, with $a_2 > a_1$:

$$I_{CSP} = (\ln(q_1/q_2))/D. \tag{5}$$

A Taylor series confidence interval is given by:

$$I_{CSP} \pm 1{\cdot}96 \times \frac{1}{D} \sqrt{[(1-q_1)/N_1 q_1 + (1-q_2)/N_2 q_2]}. \tag{6}$$

Seroconversion rate (SC). N_s women who are susceptible (seronegative) at their first antenatal test may be retested later in pregnancy or at delivery to check for seroconversion. Paired sera must be tested simultaneously, but even then the criterion for a rise in titre indicative of infection may be difficult to define. The annual infection rate is the number of seroconversions divided by the total time at risk, in years, summing over those who do not seroconvert, with a half period correction to the time at risk for those who do:

$$I_{SC} = \frac{R_{SC}}{\Sigma_i t_i + 0{\cdot}5\,\Sigma_j s_j} \tag{7}$$

where number of women initially seronegative $= N_s$; number of women seroconverting $= R_{SC}$; time between first and last negative tests in those remaining seronegative $= t_i$, $(i = 1, 2, \ldots, N_s - R_{SC})$; and time between negative and positive tests in those seroconverting $= s_j$, $(j = 1, 2, \ldots, R_{SC})$.

Where information on individual times at risk is not available, an average time at risk t can be used. This could be for example the time between the initial antenatal sample and recall 3–4 months later [19], or the time between the antenatal test and delivery [20]. Because R_{SC} is usually very small relative to the number of susceptible women at risk, the following approximation is adequate:

$$I_{SC} \simeq R_{SC}/(N_s t). \tag{8}$$

The number of seroconversions is a random event in time. Ninety-five per cent confidence limits for I_{SC} can be found by substituting 95% confidence limits for R_{SC}, based on the Poisson distribution [21], into equations (7) or (8).

Specific IgM tests for recent infection. Recent infection results in measurable levels of specific IgM. IgM tests are generally carried out on sera already known to contain specific IgG. The N_s IgG seronegatives are effectively each at risk for a period equal to the duration of IgM persistence following infection, t_p. The R_{IgM} with detectable IgM were each at risk for half that period.

$$I_{IgM} = R_{IgM}/t_p(N_s + 0{\cdot}5R_{IgM}) \simeq R_{IgM}/(N_s t_p). \tag{9}$$

https://doi.org/10.1017/S0950268800049839 Published online by Cambridge University Press

This calculation assumes that IgM is not produced in response to reinfection or recurrent infection, and it will overestimate the infection rate if this is incorrect. Approximate confidence intervals for I_{IgM} can be found by the same method given for I_{SC}. The additional uncertainty about I_{IgM} due to uncertainty in the true value of t_p should not be overlooked.

Combined programme. Some studies have used a combined programme in which initially IgG-positive sera are tested for IgM and initial negatives are retested later for seroconversion [20]. The incidence rates I_{SC} and I_{IgM} should be estimated separately and compared. If they are not statistically different [22], a single pooled estimate can be calculated:

$$I_{pooled} = (R_{SC} + R_{IgM})/N_s(t + t_p). \quad (10)$$

INCIDENCE PER 1000 PREGNANCIES

Having first estimated the annual incidence per 1000 susceptibles, the number of pregnancies in which a primary infection will occur, per 1000 consecutive pregnancies, can be calculated, taking into account that pregnancy lasts for 40 weeks, and that primary infections only occur in seronegative women.

$$\text{Incidence per pregnancy} = q(1 - \exp(-40I/52)), \quad (11)$$

where annual incidence rate $= I$, and proportion seronegative at conception $= q$. As q is usually based on large numbers, an approximate confidence interval can be constructed by substituting into (11) the confidence limits for I.

OTHER METHODS

Annual increase in seroprevalence. Some authors have calculated the 'antibody acquisition rate' or annual increase in seroprevalence [19, 20, 23]. This may be estimated by the slope H in a plot of seroprevalence against age. In the notation of equation (3):

$$1 - q(a) = Ha + k. \quad (12)$$

The rate of change of seroprevalence is not the same as the incidence, as the former takes no account of the changing number of susceptibles.

Detection rate. A detection rate is the number of primary infections per 1000 consecutive pregnancies detected by a given programme. The following formulae would cover detection by seroconversion, IgM, or both. Where N is the total number of pregnant women, seropositive and seronegative, and N_S the number who are seronegative:

$$\text{Detections per pregnancy} = (R_{SC} + R_{IgM})/N. \quad (13)$$

$$\text{Detections per susceptible pregnancy} = (R_{SC} + R_{IgM})/N_s. \quad (14)$$

Unlike the annual incidence and the incidence per pregnancy, in which the time at risk t or t_p is taken into account, the detection rate is protocol-specific. For this reason, the detection rate may not be interpreted as an infection rate per pregnancy (except in a programme where $(t + t_p = 40$ weeks).

A further complication arises in studies which include an IgM test component. IgM persistence varies considerably between individuals, and it may exceed the

time period between conception and testing. A number of the primary infections detected, and included in the detection rate, may therefore not be infections during pregnancy. If the probability distribution of IgM persistence following infection was known it would be possible to adjust formulae (13) and (14), but such information is not yet available for any IgM assay. As a result, particularly with long or highly variable t_p, the detection rate may not be a useful indicator of what an IgM or combined screening programme is achieving. Some quantitative results on the effect of the probability distribution of t_p on the sensitivity and predictive value of IgM tests in prenatal programmes are presented elsewhere [24].

The estimation of an annual incidence rate from IgM data using equations (9) and (10) is not affected by variation in persistence, and only requires assumptions regarding the mean.

RESULTS AND COMPARISON WITH PREVIOUS ESTIMATES

In the seven publications reviewed, one of which includes both CMV and toxoplasmosis studies, 18 independent estimates of annual incidence or incidence per pregnancy could have been computed, with confidence intervals. In fact, 11 independent calculations were performed, and no confidence intervals were given. Of these, two were detection rates per susceptible pregnancy [19, 25], three were detection rates per pregnancy in which no account was made of additional pre-pregnancy infections that would have been detected by IgM [6, 20], and four were annual changes in seroprevalence [19, 20, 23]. There was one incorrectly derived IgM-based infection rate per pregnancy, and one correctly derived SC-based estimate [23]. Two of three publications did not cite a mean IgM persistence for the assay used.

A wide variety of calculations have been performed, which makes it difficult to compare the results of different studies. Annual changes in seroprevalence have been interpreted as annual incidence [19, 20], and detection rates based on different protocols have not only been compared as if they were estimates of the same thing, but have been referred to as infection rates per pregnancy [6, 19, 20].

Annual incidence estimates based on CSP, SC, IgM and pooled methods [equations (4), (8), (9) and (10)] applied to the published data, are shown in Tables 1 and 2. Incidence per 1000 pregnancies is also given, based on equation (11), and using where possible the proportion seronegative from the same study.

Toxoplasmosis

For toxoplasmosis, estimates of annual incidence cover a wide range (Table 1). At one extreme is the very low SC-based estimate from the Inverness study, 1·8 per 1000 per year. This may be due to difficulty with the criterion for seroconversion. (There were 83 two-fold or more rises in titre compared to 36 falls, and yet only 4 of the rises were confirmed.) The IgM-based estimate in this study is also based on small numbers, and has broad confidence intervals. At the other extreme is the 24 per 1000 per year CSP estimate based on London data collected in the early 1970s, which is much higher than any other estimate. However, the six remaining estimates within these two extremes still show considerable heterogeneity, ranging from 4·0 to 9·5 per 1000 per year.

Table 1. *Toxoplasmosis: estimates of the proportion seronegative, annual incidence per 1000 susceptibles, and the rate of infection per 1000 pregnancies, based on material from four published UK studies*

Study [reference]	Proportion seronegative	Method	Annual incidence per 1000 (95% CI)	Rate per 1000 pregnancies (95% CI)
London [29]	0·78	CSP	24·3 (19–30)	14·5 (11–18)
		SC	4·9 (2–10)	2·9 (1–6)
Glasgow [19]	0·87	SC	7·8 (5–12)	5·2 (3–8)
W. Scotland [19]	0·86	CSP	5·0 (4–6)	3·3 (2–4)
E. Scotland [19]	0·75	CSP	9·5 (6–13)	5·5 (3–8)
Inverness [20]	0·83	CSP	6·3 (4–8)	4·0 (3–5)
		SC*	1·8 (0·5–4)	1·1 (0–3)
		IgM 3 m†	6·1 (2–13)	3·9 (1–9)
		IgM 6 m	3·0 (1–7)	1·9 (0–4)
		IgM 9 m	2·0 (1–4)	1·3 (0–3)
W. Glamorgan [6]		IgM 7·5 m‡	4·0 (3–5)	§2·5 (2–3)

* Based on 4748 paired sera, of which 83·1% were initially seronegative for toxoplasmosis (Table 2, ref. [20]).

† No mean IgM persistence value available: incidence estimated on a range of assumptions.

‡ IgM persistence was between 6 and 9 months (personal communication).

§ Assuming 81% seronegative, the average of the studies cited.

Table 2. *Cytomegalovirus: estimates of the proportion seronegative, annual incidence per 1000 susceptibles, and the rate of infection per 1000 pregnancies, based on material from four published UK studies*

Study [reference]	Proportion seronegative	Method	Annual incidence per 1000 (95% CI)	Rate per 1000 pregnancies (95% CI)
Edinburgh [25]	0·46	SC	13·1 (6–26)	4·6 (2–9)
Inverness [20]	0·59	CSP	19·2 (15–23)	8·7 (7–11)
		SC*	8·2 (4–14)	3·6 (2–6)
		IgM 3 m	14·5 (7–27)	6·6 (3–12)
		IgM 6 m	7·2 (3–13)	3·3 (2–6)
		IgM 9 m	4·8 (2–9)	2·2 (1–4)
London [23]	0·42	CSP†	23·7 (19–28)	7·6 (6–9)
		SC	17·9 (12–25)	5·7 (4–8)
		IgM 4 m	28·9 (16–47)	9·2 (5–15)
		Pooled	20·5 (13–29)	6·6 (4–9)
Cardiff [30]	0·46‡	CSP	37·7 (29–47)	13·2 (10–16)

* Based on 4717 sera of which 59% were initially seronegative (Table 2, ref. [20]).

† From data kindly supplied by Professor P. Griffiths.

‡ Average of percent seronegative aged 17–44, weighted to reflect age distribution of pregnant women.

The proportion of toxoplasmosis seronegatives in the studies reviewed ranged from 75–86%. Estimates of the number of pregnancies per 1000 in which a maternal infection occurs ranged from 2·5–5·5, based on the six central estimates.

It may be significant that three of the four highest estimates are CSP based. Two recent studies of sera collected over a 20- or 30-year period have shown that age-specific seroprevalence has been declining [26, 27]. Under these circumstances,

the CSP method, which in antenatal sera effectively reflects incidence 20–40 years ago, would yield overestimates.

Cytomegalovirus infection

The annual incidence estimates for CMV infection also covered a wide range (Table 2). Assuming CMV-specific IgM persistence to have been under 6 months in the Inverness study, the estimates of annual incidence range from 8·2–37·7 per 1000 susceptibles. The overall percentage of seronegatives ranges from 42–58%. The frequency with which a primary infection occurs in pregnancy is in the range 3·7–13·4 per 1000, with most studies between 4 and 10.

DISCUSSION

This review of the published findings on the incidence of toxoplasmosis and CMV infection in pregnancy has shown that previous investigators have not made the best use of their data. A variety of estimates were derived, most of which are protocol specific. Detection rates per pregnancy or detection rates per susceptible pregnancy have been cited as 'infection rates' or 'rates per pregnancy' and compared with each other, without taking account of periods at risk. The extent to which IgM tests may identify pre-pregnancy infection [24] has not been allowed for, and the implications of IgM persistence for measurements of incidence has been widely overlooked. Only one incidence rate per pregnancy was correctly derived.

In view of the variety of calculations performed, there is surprising unanimity in the widely cited 2 per 1000 pregnancies estimate of toxoplasmosis incidence [1, 4, 6, 7, 19, 20]. However, when reanalysed, the available data yield a range of incidence estimates between 1·1 and 14·4 per 1000 pregnancies, with most between 2·5 and 5·5 per 1000. A discussion of possible behavioural or geographical reasons for this heterogeneity would be beyond the scope of this paper. However, the results show that it may not be safe to base expectations for proposed screening programmes on research carried out in another time or place.

None of the published estimates have been accompanied by confidence intervals. Only two studies had more than 25 seroconversions and/or IgM positives. With this sample size, the upper 95% confidence limit is over two times greater than the lower. Some of the debate on screening has been informed by estimates based on 10 infections, whose confidence intervals have a 3·8-fold span.

It may finally be useful to briefly mention the likely effect on the potential role of prenatal screening of altering assumptions about the incidence of infection in pregnancy.

In the case of CMV infection, the incidence of congenital infection in the UK and the percent distribution of its sequelae have been well characterised by follow-up studies of children born to women seroconverting in pregnancy [9, 23, 28]. The present range of estimates, 4–10 per 1000 pregnancies, is consistent with accepted figures and does not alter arguments for screening.

For toxoplasmosis, however, neither the number of congenital infections per year, nor the precise distribution of their sequelae is sufficiently known [1]. Under these circumstances the arguments for or against screening depend partly on what

one assumes the true maternal incidence to be, and partly on assumptions about the sequelae of maternal infection. If all symptomatic cases of congenital toxoplasmosis are already recognized, but if incidence in pregnancy is twice what was thought, then screening would lead to twice as many unnecessary interventions. However, if serious symptomatic illness is occurring but not being recognized as resulting from congenital infection, then the higher incidence in pregnancy would indicate that screening may have a greater potential to prevent serious morbidity.

Given the immense effort required to follow thousands of women through pregnancy, and to carry out laboratory tests for extremely rare conditions, it is unfortunate that the data have not always been fully exploited. This is data on which decisions with far-reaching financial and human implications will be based. Toxoplasmosis antenatal screening programmes are now under way in several major maternity centres in the UK. It is to be hoped that appropriate statistical methods will be used to analyse the results, so that programmes can be compared and an overall evaluation of the benefits of screening can be made.

REFERENCES

1. Editorial. Antenatal screening for toxoplasmosis in the UK. Lancet 1990; ii: 346–8.
2. Ho-Yen DO, Chatterton JMW, Joss AWL. Screening for infection in pregnancy. Lancet 1988; ii: 1031.
3. Ho-Yen DO. Maternal and fetal screening. Br M J 1990; **300**: 1527.
4. Joss AWL, Chatterton JMW, Ho-Yen DO. Congenital toxoplasmosis: to screen or not to screen? Public Health 1990; 104: 9–20.
5. Ho-Yen DO, Joss AWL. Toxoplasma and cytomegalovirus infections during pregnancy. Matern Child Health 1988; **13**: 225–7.
6. Joynson DHM, Payne R. Screening for toxoplasma in pregnancy. Lancet 1988; ii: 795–6.
7. Ho-Yen DO. Toxoplasmosis in humans: discussion paper. J R Soc Med 1990; **83**: 571.
8. Ho-Yen DO, Chatterton JMW. Congenital toxoplasmosis – why and how to screen. Med Microbiol 1990; **1**: 229–35.
9. Preece P, Pearl K, Peckham C. Congenital cytomegalovirus. Arch Dis Childhood 1984; **59**: 1120–6.
10. Hennekens CH, Buring JE. Epidemiology in medicine. Boston/Toronto: Little, Brown. 1989.
11. Anderson RM, May RM. Vaccination against rubella and measles: quantitative investigations of different policies. J Hyg 1983; **90**: 259–325.
12. Knox EG. Strategy for rubella vaccination. Int J Epidemiol 1980; **9**: 13–23.
13. van Druten H, van Knapen F, Reintjes A. Epidemiologic implications of limited-duration seropositivity after toxoplasma infection. Am J Epidemiol 1990; **132**: 169–80.
14. Griffiths DA. A catalytic model of infection for measles. Appl Stat 1974; **23**: 330–9.
15. Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. J Hyg 1985; **95**: 419–36.
16. Becker NG. Analysis of infectious disease data. New York: Chapman and Hall. 1989.
17. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol 1986; **123**: 174–84.
18. Ades AE, Peckham CS, Dale GE, Best JM, Jeansson S. Prevalence of antibodies to herpes simplex virus types 1 and 2 in pregnant women, and estimated rates of infection. J Epidemiol Community Health 1989; **43**: 53–60.
19. Williams KAB, Scott JM, MacFarlane DE, Williamson JMW, Elias-Jones TF, Williams H. Congenital toxoplasmosis: a prospective survey in the West of Scotland. J Infect 1981; **3**: 219–29.
20. Joss AWL, Skinner LJ, Chatterton JMW. Simultaneous serological screening for congenital cytomegalovirus and toxoplasma infection. Public Health 1988; **102**: 409–17.

21. Lentner C. (ed.) Geigy scientific tables, vol 2. Geneva: Ciba-Geigy, 1982.
22. Rothman KJ. Modern epidemiology. Boston/Toronto: Little, Brown, 1986: 358.
23. Griffiths PD, Baboonian C. A prospective study of primary cytomegalovirus infection during pregnancy: final report. Br J Obstet Gynaeco 1984; **91**: 307–15.
24. Ades AE. Evaluating the sensitivity and predictive value of tests of recent infection: toxoplasmosis in pregnancy. Epidemiol Infect 1991; **107**: 527–35.
25. Grant S, Edmond E, Syme J. A prospective study of primary cytomegalovirus infection in pregnancy I: Laboratory evidence of congenital infection following maternal primary and reactivated infection. J Infect 1983; **3**: 24–31.
26. Forsgren N, Gille E, Ljungstrom I, Nokes DJ. *Toxoplasma gondii* antibodies in pregnant women in Stockholm, 1969, 1979, and 1987. Lancet 1991; **337**: 1413–4.
27. Walker J, Nokes DJ, Jennings R. Longitudinal study of toxoplasma seroprevalence in South Yorkshire. Epidemiol Infect. In press.
28. Pearl K, Preece P, Ades A, Peckham C. Neurodevelopmental assessment after congenital cytomegalovirus infection. Arch Dis Child 1986; **61**: 323–6.
29. Ruoss CF, Bourne GL. Toxoplasmosis in pregnancy. J Obstet Gynaecol 1972; **79**: 1115–8.
30. Yarnell JWG, Milbank JE. The prevalence of cytomegalovirus antibody in women: an epidemiological study from south Wales. Public Health London 1982; **96**: 251–5.