Epidemiol. Infect. (1991), **107**, 527–535
Printed in Great Britain

Evaluating the sensitivity and predictive value of tests of recent infection: toxoplasmosis in pregnancy

A. E. ADES

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

(*Accepted 14 May 1991*)

SUMMARY

The diagnosis of maternal infection in early pregnancy depends on tests which are sensitive to recent infection, such as specific IgM. Two types of test are considered: those where the response persists for a period following infection and then declines, such as IgM, and those whose response increases with time since infection, such as IgG-avidity. However, individuals vary in their response to infection, and it may not always be possible to determine whether an infection occurred during pregnancy or before it. Mathematical methods are developed to evaluate the performance of these tests, and are applied to the diagnosis of toxoplasmosis in pregnancy. It is shown that, based on existing information, tests of recent infection are unlikely to be both sensitive and predictive. More data on these tests are required, before they can be reliably used to determine whether infection has occurred during pregnancy or before it.

INTRODUCTION

The potential advantages of screening for toxoplasmosis in pregnancy, in order to prevent congenital toxoplasmosis, have been widely debated in several countries [1–7]. The foetus is only at risk following a primary maternal infection during pregnancy. Methods for the serodiagnosis of primary infection have been reviewed recently [6, 8]. While the presence of toxoplasma-specific IgG on recall following an initial negative test can provide evidence of seroconversion in pregnancy, an infection that occurs after conception but before the first antenatal visit can only be identified by tests sensitive to recent infection, such as toxoplasma-specific IgM. The use of such tests in prenatal screening programmes has been recommended by several experts [2, 9]. Tests for IgG-avidity [10] or toxoplasma-specific IgA [11], which may also be able to discriminate recent from past infection, have also been proposed.

Similar serodiagnostic methods can be used to diagnose other primary infections which threaten the foetus. Specific IgM tests are used routinely to diagnose maternal rubella infection following a contact or a suspected contact in pregnancy. They have a similar role in cytomegalovirus (CMV), and have also been considered in the context of antenatal screening programmes for CMV [12].

Even if a test is 100% sensitive and 100% specific in identifying women with recent infection, it may not be possible to determine whether a recent infection

occurred during pregnancy or before it, because of individual variation in response to infection. In this paper this issue is explored mathematically. Simple assumptions are made about the degree of between-patient variation, based on the best information available. It is then possible to calculate the sensitivity and predictive value that can be achieved when tests of recent infection are interpreted as tests of infection during, as opposed to before, pregnancy.

METHODS AND RESULTS

Taking IgM and IgG-avidity tests in turn, this section develops an analysis of the testing situation and provides illustrative results, based where possible on published data on tests for recent toxoplasmosis infection. Mathematical details can be found in an Appendix.

Interpretation of IgM tests

Following a primary infection, IgM increases and then declines. The duration of its persistence depends on the peak level reached and the rate of decay, which will both vary among individuals. However, it is instructive to assume initially that there is no such variation. For example, if IgM always persists for exactly 20 weeks following infection, then all infections in pregnancy in women tested at 12 weeks gestation would be detected (100% sensitivity), but so would many infections that predate pregnancy. The probability that a positive test resulted from an infection that occurred during pregnancy, the positive predictive value (PPV), would be 60% (12/20). Similarly, if IgM persisted for exactly 8 weeks in all individuals, then its presence at 12 weeks gestation would confirm an infection during pregnancy (100% Positive Predictive Value, PPV), but would exclude those infected in the first 4 weeks. The effective sensitivity would only be 67% (8/12).

One IgM test that has been used to assess the incidence of toxoplasmosis in pregnancy [2] is thought to have mean persistence of between 6 and 9 months (personal communication). The PPV for women tested at 18 weeks would therefore be between 69% and 46%, so that between 31% and 54% of those who are positive on this test would in fact have become infected before their pregnancy. They would therefore risk going on to have unnecessary treatment, further investigation, or termination.

When the persistence of the IgM response to infection does not match up very closely indeed with the time between conception and testing, then the test will either become insensitive or have poor PPV. In theory, a test with 12-week persistence would give perfect results for a woman tested at 12 weeks. However, this assumes that there is no variation between individuals in IgM persistence. If assumptions are made about the distribution of IgM persistence it is possible to calculate the effect of between-patient variation on sensitivity and PPV.

It is convenient and plausible to assume that persistence has a lognormal distribution. The spread of the distribution is expressed as a geometric standard deviation (GSD). In a test with a median persistence of 16 weeks, a GSD of 1·4 would mean that 95% of infected individuals would have a persistence between 8·3 and 30·9 weeks (calculated from $16/(1{\cdot}4^{1{\cdot}96})$ and $16(1{\cdot}4^{1{\cdot}96})$). Table 1 column 2 shows how the 95% limits vary with GSD.

https://doi.org/10.1017/S0950268800049220 Published online by Cambridge University Press

Table 1. *Sensitivity and Positive Predictive Value of IgM tests for infection during pregnancy, as a function of the variation in persistence between individuals, expressed as the geometric SD. (3): sensitivity and PPV when the median persistence exactly equals time between conception and testing. (4): maximum PPV attainable with 95% sensitivity. (5): maximum PPV attainable with 98% sensitivity*

(1) Geometric SD	(2) 95% range of persistences*	(3) Sensitivity (%) and PPV (%)	(4) PPV (%) at SE = 95%	(5) PPV (%) at SE = 98%
1·2	11·2–22·9	94	90	85
1·3	9·6–26·8	91	85	79
1·4	8·3–30·9	89	80	71
1·6	6·4–35·4	86	70	59
1·8	5·1–50·6	83	63	50
2·0	4·1–62·3	81	57	45

* Assuming a 16 week median persistence.

Consider, firstly, a situation where the median IgM persistence following infection is equal to the gestational age at testing. As a result of the lognormal assumption, sensitivity and PPV are equal under these circumstances. For example, the proportion of infected women with IgM persistence 1·5 times more than the median (with sensitivity 100% and PPV 67%) will be the same as the proportion with persistence 1·5 times less (with sensitivity 67% and PPV 100%). Table 1 column 3 gives the sensitivity and PPV obtained as GSD increases. They fall to below 90% at a GSD of 1·4.

There will usually be both between-patient variation in IgM persistence and a mismatch between median persistence and gestational age at the test. Table 1 also shows the best PPV that could be obtained in tests with 95% sensitivity (column 4), or 98% sensitivity (column 5). These PPVs are theoretical in the sense that they assume that IgM tests can be chosen with the median persistence that will give the best results at a given gestational date. The rapid deterioration in PPV occurs because, in order to achieve reasonable sensitivity in the face of even moderate between-patient variation, the median IgM persistence has to be considerably longer than gestational age at testing, which lowers PPV. For example, to achieve 95% sensitivity at a GSD of 1·6, the median persistence of the test would have to be approximately 1·43 times the gestational age at testing, yielding a PPV of 70%.

Test response related to time since infection: IgG-avidity

A more precise diagnosis may be possible in a test such as IgG-avidity, which gives a continuous response that can be related, in a more or less linear fashion, to time since infection [10]. A regression of test response against time since infection is shown in Fig. 1. Such a graph would be based on repeat testing of individuals infected at a known date. A woman's test response can then be compared to the response expected from someone infected since the beginning of her pregnancy. While this solves the problem of tests having a fixed median persistence, which may not match up with the gestational age of testing, it does

https://doi.org/10.1017/S0950268800049220 Published online by Cambridge University Press

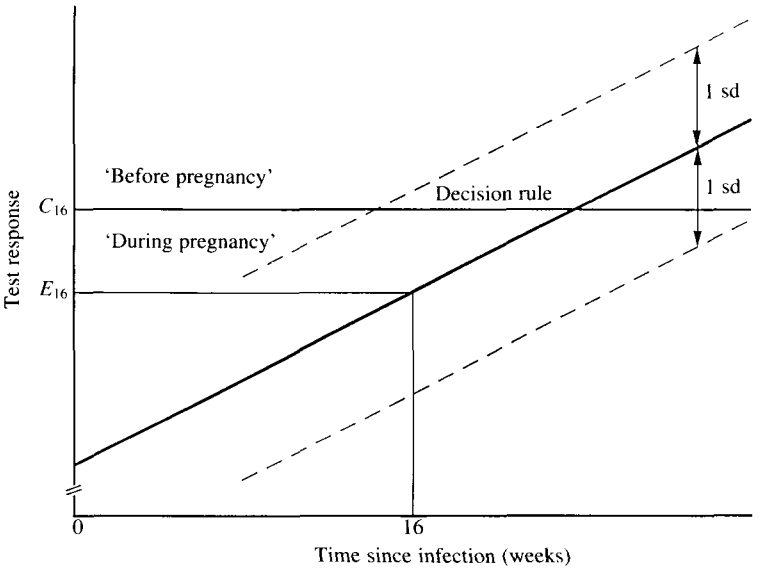

Fig. 1. A regression of test response against time since infection, showing the between-patient variation in test response about the line. For a woman tested at 16 weeks gestation, the most likely test response is E_{16}. The line C_{16} is the cut-off for the decision rule: assume that infection occurred *during* pregnancy if the observed test result is less than C_{16}, and *before* pregnancy if it is not. The graph is drawn to a scale in which the SD:slope ratio is 6 weeks per SD unit of response.

not solve the problem of between-patient variation, which is depicted in Fig. 1 by the two lines placed one standard deviation (SD) unit above and below the regression line.

Using the regression line and the SD about it, it is possible, given a test result obtained at a known gestational date, to calculate the probability that the infection had occurred during the pregnancy. In order to assess the overall performance of this test system, we assume that a decision as to whether the infection occurred during or before pregnancy will be based on a simple rule. If the test response is higher than some preset cut-off, the decision is *before*, and if it is lower the decision is *during*. This is shown in Fig. 1. The most appropriate position of the decision rule cut-off will depend on the gestational age at testing. In Fig. 1 the cut-off C_{16}, for women tested at 16 weeks gestation, is set at a point where a woman infected 16 weeks earlier would be likely to give a test response below the cut-off, leading to a correct 'during' decision. However, it is clear that a proportion of women infected 17, 18 or 19 weeks earlier, before pregnancy, would also give rise to responses that would lead to a 'during' decision. These are false positives. At the same time some women infected only 14 or 15 weeks earlier may have test responses above C_{16}, leading to false negative before decisions. The cut-off can be lowered to reduce the false positives, raising the PPV. But this can only be done at the expense of increasing the number of false negatives, lowering sensitivity.

The overall resolving power of the test system at a given gestational age, its ability to discriminate infection during pregnancy from infection before pregnancy, is controlled by a single parameter, the ratio of the SD to the slope. For

https://doi.org/10.1017/S0950268800049220 Published online by Cambridge University Press

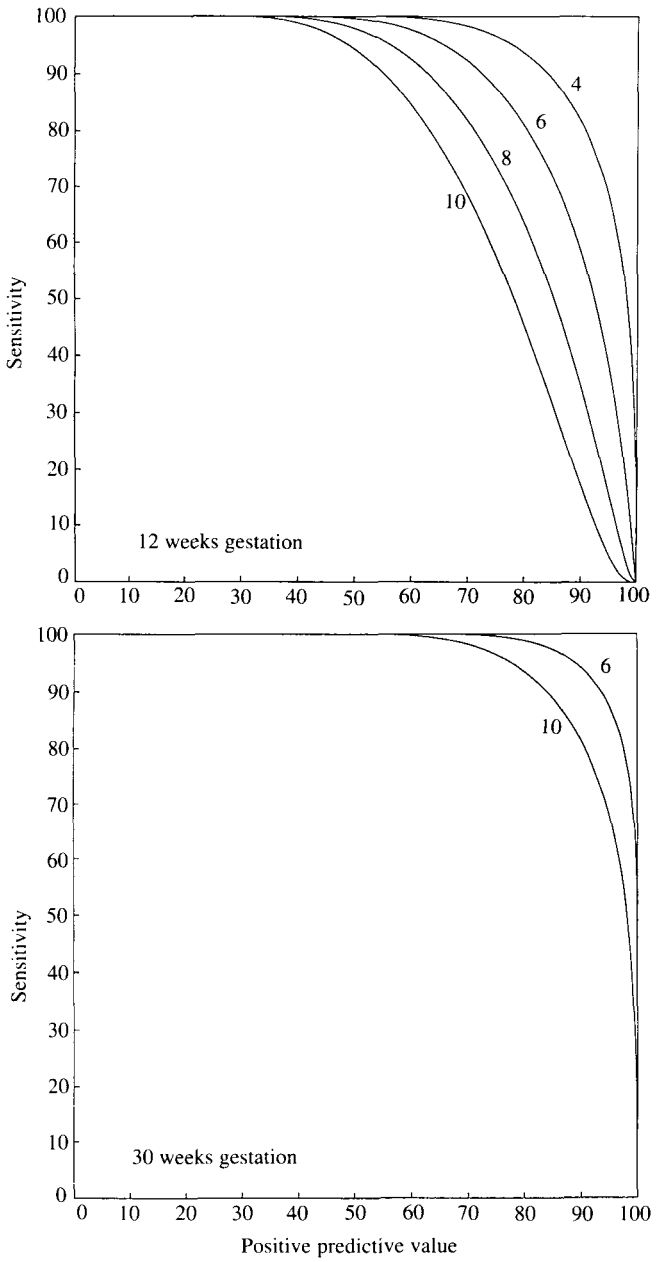

Fig. 2. Sensitivity plotted against Positive Predictive Value, with different values of the SD:slope ratio. Each curve is generated by moving the decision rule cut-off from very low to very high. (a) women tested at 12 weeks gestation, (b) 30 weeks gestation.

a given sensitivity, one can obtain improved PPV by increasing the slope, or by narrowing the SD.

In Fig. 2 sensitivity is plotted against PPV for women tested (a) at 12 weeks and (b) at 30 weeks gestation. Each curve is generated by moving the decision rule cut-off from very low (low sensitivity, high PPV) to very high (high sensitivity, low

https://doi.org/10.1017/S0950268800049220 Published online by Cambridge University Press

PPV). The SD:slope ratio is varied to show how either a low slope or a high SD will degrade the test resolution. For example, for women tested at 12 weeks, if the test system has an SD:slope ratio of 6, a decision rule that correctly picks out 95% of all those infected in pregnancy, will have a predictive value of only 65%, so that 35% of those picked out will have been infected before their pregnancy. Alternatively, the obstetrician could choose a decision rule with a 90% PPV: but this would have a sensitivity of only 59%, so that this rule would fail to identify 41% of women infected in pregnancy. The only way to improve both sensitivity and PPV simultaneously is to lower the SD:slope ratio. It should be noted, however, that the SD:slope ratio is an inherent property of the test, and cannot be changed without changing the test.

Fig. 2 also shows how the resolution of the test increases markedly for women tested later in pregnancy. The reason for this is that at 30 weeks gestation, for example, half the women infected during pregnancy would have been infected over 15 weeks earlier, all of whom would be very likely to be detected even if the cut-off is set low to improve PPV. For women at 30 weeks gestation, a test system with an SD:slope ratio of 6, could obtain a sensitivity of 95% with 88% PPV, compared with 65% PPV in women tested at 12 weeks gestation.

No estimates of the SD:slope ratio have been published for any test system. However, in the case of IgG-avidity, it is possible to derive an estimation from published data [10, (Table 1)]. A log transform was applied to the percent avidity data, as this produced a better fitting regression line. The slope was 0·095 log units per week (95% confidence interval: 0·048 − 0·14) with a residual SD of 0·61 about the line, giving a SD:slope ratio of 6·4, close to the scale used in Figure 1. However, this small dataset, with only 17 data points from 10 patients tested between 2 and 26 weeks gestation, would be compatible with SD:slope ratios which are higher or lower by a factor of at least 2. In addition the data are insufficient to test the assumption that the test response is linear, and that the SD is constant, or can be made constant by transformation, over time since infection. If the variation about the line increased with time since the infection, then the dramatic improvement in resolution at later gestational ages (Fig. 2b) would not occur.

DISCUSSION

Proponents of prenatal screening for toxoplasmosis and other infectious disease threatening the foetus have undoubtedly been aware that IgM persistence varies between individuals. One objection to these tests has been their lack of specificity as screening tests for a low prevalence condition [1, 7]. What may have been lacking is a discussion of how they would perform as diagnostic tests of infection during as opposed to before pregnancy, even assuming 100% sensitivity and specificity in a biological sense.

Our results show that irrespective of between-patient variation IgM tests will be either sensitive or predictive, but not both. There is no published data to suggest that the persistence of any IgM test is accurately known, and no one has ventured to give an estimate of between-patient variation in IgM persistence. As there are many anecdotal reports of IgM persisting for 1, 2 or even 3 years, it would be optimistic to believe that the GSD was less than 1·5. At this level, 75% is the

theoretical maximum PPV that could be achieved with 95% sensitivity, assuming that an IgM test with appropriate median persistence was available. Whether or not this is regarded as a satisfactory level of performance, median IgM persistence and between-patient variation must first be reliably quantified before such tests are included in a programme of screening and diagnosis.

Tests giving a continuous response, which varies linearly with time since infection, while at an earlier stage of development, have been regarded as more promising [6]. However, in order to demonstrate their usefulness, it is not enough to show that the mean response of those infected recently is different from those infected in the past [10]. Large-scale studies are required on patients whose primary toxoplasmosis infection has a known date of onset, with repeat samples every month for at least a year. The slope of the response/time curve, and the variation around it, can then be reliably assessed. On the data available at present 95% sensitivity could only be achieved with 65% PPV.

Although new tests of recent infection are likely to be proposed in the future, it seems inevitable that they will either have a response like IgM, which persists for a period after infection, or a response that can be related to time since infection. Between-patient variation is certain to be a feature of any new test. It is therefore likely that the issues addressed here will apply to any test of recent infection that is used to assist the diagnosis of primary infection in pregnancy. The methods used to evaluate test reliability in this setting are not restricted to toxoplasmosis. They could be applied to other primary infections that carry a risk of foetal damage and in which the date of infection cannot always be determined, such as cytomegalovirus, rubella, and parvovirus B19. Such evaluations are becoming increasingly necessary because of the very high sensitivity of recently available IgM tests.

In practice, the final clinical decision is based on the results of several tests carried out possibly over a period of weeks, as well as the clinical picture. While quantitative statements about how all the information is put together cannot easily be made, this does not alter the need for quantitative evaluation of laboratory tests. Furthermore, a number of simplifying assumptions have been made in the calculations presented here, which have been conservative, in the sense that the performance of tests of recent infection would be worse in real life. For example, we have assumed 100% sensitivity and specificity at a biological level, whereas IgM tests in one proposed antenatal screening programme were less than 95% specific [9]. Similarly, it has been assumed that the objective is to decide whether infection occurred before or during infection. However, an infection acquired just before conception may lead to a parasitaemia which lasts into pregnancy, and which is capable of infecting the fetus. This not only shifts the goalpost for tests of recent infection, but turns it into a moving one, as the duration of the parasitaemia will also vary between patients.

While there is an urgent need for more serious work on the effective sensitivity and PPV of test systems for recent infection, it is less clear how much sensitivity and predictive value is required. This will depend on whether the test result is to be the triggering event for a decision to terminate, or to treat, or to attempt a definitive diagnosis of foetal infection. To some extent this will also depend on the gestational age at the first antenatal test. The current medico-legal environment

appears to put considerable weight on false negative diagnoses, thereby favouring high sensitivity at the cost of low predictive value. The consequence would be a high number of unnecessary treatments, further investigations and terminations of pregnancy. In the absence of clear guidelines based on known risks and benefits of these procedures, and on the known properties of the test systems, practice may differ very widely, depending only on obstetricians' personal views.

APPENDIX

IgM tests with no between-patient variation in persistence

For a woman tested at G weeks gestation on a test with persistence P, the sensitivity SE and positive predictive value PPV are:

$$SE(G;P) = P/G \text{ if } P < G, \text{ otherwise } 1$$

$$PPV(G;P) = 1 \text{ if } P < G, \text{ otherwise } G/P$$

IgM tests with between-patient variation in persistence

Assume that persistence P is lognormally distributed with median M, and geometric SD S, with density $LN(P;M,S)$. Sensitivity and PPV of the test in a woman tested at G weeks gestation is:

$$SE(G;M,S) = \int_0^\infty SE(G;P)\,LN(P;M,S)\,dP$$

$$PPV(G;M,S) = \int_0^\infty PPV(G;P)\,LN(P;M,S)\,dP$$

These are the values tabulated in Table 1, with different values of S. In Column (3) $G = M$, while in columns (4) and (5) SE and PPV were generated by varying M at a fixed G and interpolating.

Test response linearly related to time since infection (IgG avidity)

Test response y is normally distributed, with between-patient standard deviation S about a mean value that is linearly related to time since infection w. Represent the density as $N(y;b_0+b_1w,S)$, where b_0 and b_1 are constants.

In a test carried out at G weeks gestation, the probability that $y < C_G$ in a woman infected between t_1 and t_2 weeks after conception can be written as:

$$F(G,C_G/t_1 < w < t_2) = \frac{1}{(t_2-t_1)}\int_{t_1}^{t_2}\int_{-\infty}^{C_G} N(y;b_0+b_1w,S)\,dy\,dw$$

Then, with a decision rule: *DURING pregnancy if* $y < C_G$, *otherwise BEFORE pregnancy*, the sensitivity and PPV for a woman tested at G weeks gestation is:

$$SE(G;C_G) = F(G,C_G/0 < w < G)$$

$$PPV(G;C_G) = \frac{G^*F(G,C_G/0 < w < G)}{G^*F(G,C_G/0 < w < G)+(k-G)^*F(G,C_G/G < w < k)}$$

in the limit $k \to$ infinity

https://doi.org/10.1017/S0950268800049220 Published online by Cambridge University Press

In Fig. 2, a SE/PPV curve is generated by varying C_G. This is done for $G = 12$ and $G = 30$, and for various values of S/b_1. Calculations were carried out with FORTRAN programmes using NAG library subroutines.

ACKNOWLEDGEMENTS

The author thanks David Winstanley of the University of London Computing Centre for his assistance in the use of the NAG subroutine library.

REFERENCES

1. Editorial. Antenatal screening for toxoplasmosis in the UK. Lancet 1990; **2**: 346–8.
2. Joynson DHM, Payne R. Screening for toxoplasma in pregnancy. Lancet 1988; **2**: 795–6.
3. Ho-Yen DO. Toxoplasmosis in humans: discussion paper. J R Soc Med 1990; **83**: 571.
4. Ho-Yen DO, Chatterton JMW. Congenital toxoplasmosis – why and how to screen. Med Microbiol 1990; **1**: 229–35.
5. McCabe R, Remington JS. Toxoplasmosis: the time has come. N Engl J Med 1988; **318**: 313–5.
6. Koskiniemi M, Lappalainen M, Hedman K. Toxoplasmosis needs evaluation. Am J Dis Child 1989; **143**: 724–8.
7. Frenkel JK. Diagnosis, incidence, and prevention of congenital toxoplasmosis. Am J Dis Child 1990; **144**: 956–7.
8. Holliman RE. The diagnosis of toxoplasmosis. Serodiag Immunother Infect Dis 1990; **4**: 83–93.
9. Joss AWL, Chatterton JMW, Ho-Yen DO. Congenital toxoplasmosis: to screen or not to screen? Public Health 1990; **104**: 9–20.
10. Hedman K, Lappalainen M, Seppala I, Makela O. Recent primary toxoplasma infection indicated by a low avidity of specific IgG. J Infect Dis 1989; **159**: 736–40.
11. Stepick-Biek P, Thulliez P, Araujo FG, Remington JS. IgA antibodies for diagnosis of acute congenital and acquired toxoplasmosis. J Infect Dis 1990; **162**: 270–3.
12. Joss AWL, Skinner LJ, Chatterton JMW. Simultaneous serological screening for congenital cytomegalovirus and toxoplasma infection. Public Health 1988; **102**: 409–17.