BJOG: an International Journal of Obstetrics and Gynaecology
June 2004, Vol. 111, pp. 521–531
DOI: 10.1111/j.1471-0528.2004.00193.x

REVIEW

SURUSS in perspective

Background Until the publication of the Serum Urine and Ultrasound Screening Study (SURUSS) report, it was difficult to compare the different antenatal screening tests for Down's Syndrome because of variations in study designs. We here present the main results from SURUSS, updated to take account of recent information on nuchal translucency in Down's Syndrome pregnancies, and discuss their implications.

Methods SURUSS was a prospective study of 47,053 singleton pregnancies (including 101 pregnancies with Down's Syndrome) conducted in 25 maternity units. Nuchal translucency measurements were taken. Serum and urine samples collected between 9 and 13 weeks, and again between 14 and 20 weeks of pregnancy were stored. Samples from each affected pregnancy and five matched controls were tested for currently used or suggested biochemical Down's Syndrome screening markers. Pregnancies were followed up to determine the presence or absence of Down's Syndrome. For an 85% Down's Syndrome detection rate, the false-positive rate for the Integrated test (nuchal translucency and pregnancy associated plasma protein-A [PAPP-A] at 11 completed weeks of pregnancy, and α-fetoprotein, unconjugated oestriol [uE_3], free β or total human chorionic gondaotrophin (hCG) and inhibin-A in the early second trimester) was 0.9%, the Serum integrated test (without nuchal translucency) 2.7%, the Combined test (nuchal translucency with free β-hCG and PAPP-A at 11 weeks) 4.3%, the Quadruple test (α-fetoprotein, uE_3, free β or total hCG and inhibin-A) 6.2%, and nuchal translucency at 11 weeks, 15.2%. All tests included maternal age. Using the Integrated test at an 85% detection rate, there would be six diagnostic procedure-related unaffected fetal losses following amniocentesis per 100,000 women screened compared with 35 using the Combined test or 45 with the Quadruple test.

Conclusions The Integrated test offers the most effective and safe method of screening for women who attend in the first trimester. The next best test is the Serum integrated test. The Quadruple test is the best test for women who first attend in the second trimester. There is no justification for retaining the Double (α-fetoprotein and hCG) or Triple (α-fetoprotein, uE_3, and hCG) tests, or nuchal translucency alone (with or without maternal age) in antenatal screening for Down's Syndrome.

Introduction

Antenatal screening for Down's Syndrome has developed over the last 15 years, with many tests now available. The Serum Urine and Ultrasound Screening Study (SURUSS) conducted between 1995 and 2002 provides the largest data set reported on women seen in both the first and second trimesters of pregnancy without intervention in the first trimester. This allows a direct and unbiased examination of the performance of first and second trimester Down's Syndrome markers used in different combinations to determine those that yield the most discriminatory, safe and cost-effective methods of screening. The full SURUSS report[1] provides the detailed results. Here we summarise the main findings and discuss advantages of the study design adopted, compare the results with some other studies and consider the practical implications arising from the results.

BOX 1	Definitions of Screening Tests
Double test	Early second trimester (14–20 completed weeks) test based on the measurement of α-fetoprotein and human chorionic gonadotrophin (hCG) (either free β-hCG or total hCG), together with maternal age.
Triple test	Early second trimester (14–20 completed weeks) test based on the measurement of α-fetoprotein, unconjugated oestriol (uE_3), and hCG (either total hCG or free β-hCG) together with maternal age.
Quadruple test	Early second trimester (14–20 completed weeks) test based on the measurement of α-fetoprotein, uE_3, free β-hCG (or total hCG) and inhibin-A together with maternal age.
Combined test	Late first trimester (10–13 completed weeks) test based on combining nuchal translucency measurement with free β-hCG, pregnancy associated plasma protein A (PAPP-A) and maternal age.

BOX 1	Definitions of Screening Tests
Integrated test	The integration of different screening markers measured at different stages of pregnancy into a single test result. Unless otherwise qualified, 'Integrated test' refers to the integration of nuchal translucency measurement and PAPP-A in the first trimester with the Quadruple test in the second.
Serum Integrated test	A variant of the Integrated test using serum markers only (PAPP-A in the first trimester and the Quadruple test in the second trimester).

Methods

SURUSS was based on 47,507 women recruited between September 1996 and April 2000 who attended 25 maternity centres (24 in the United Kingdom and one in Austria). Follow up was carried out to 31 May 2001. Outcomes were known for 96% of all pregnancies and ascertainment of Down's Syndrome was probably complete through cross-checking with the National Down's Syndrome Cytogenetic Register. There were 47,053 singleton pregnancies, including 101 with Down's Syndrome. The number of Down's Syndrome cases found was similar to that expected from the age distribution of the women recruited. The main screening tests and abbreviations for the screening markers are defined in Box 1.

Women were invited to join the study when they attended their first antenatal visit (between 9 and 13 completed weeks of pregnancy). As part of their ultrasound dating scan, participants had a nuchal translucency measurement, which was not reported or used in managing the pregnancy. At the visit, sonographers aimed to obtain at least three nuchal translucency measurements per fetus and the average was used. A blood sample and a urine sample were collected and stored at −40°C. Women provided another blood sample and urine sample in the second trimester (between 14 and 22 weeks), usually as part of the routine screening programme offered by the maternity centre. This design allowed comparison of first and second trimester screening tests without the bias caused by the diagnosis and termination of some Down's Syndrome pregnancies detected in the first trimester, and the miscarriage of others between the first and second trimester.

A nested case–control study within the cohort of 47,053 singleton pregnancies was used to estimate screening performance. Pregnancies with Down's Syndrome (cases) were matched with five singleton unaffected pregnancies (controls) by centre, maternal age, crown–rump length or a biparietal diameter if a crown–rump length was not available, and duration of sample storage. The serum samples were assayed for PAPP-A, free β-hCG, total hCG, uE_3 and PAPP-A and urine samples for invasive trophoblast antigen, β-core fragment, total hCG and free β-hCG. Urine marker levels were creatinine-adjusted to allow for urine dilution. Samples from cases and controls were assayed blind in the same analytical batch.

To allow for systematic changes in marker levels with increasing gestational age, all concentrations were converted into multiples of the normal median (MoM) for a given gestational age (or crown–rump length measurement in the case of nuchal translucency). Log transformations were applied to MoM values to produce distributions that were approximately Gaussian (except for urine total hCG, for which a square root transformation was used). These distributions in Down's Syndrome and unaffected pregnancies were specified in terms of the following parameters: means, standard deviations and correlation coefficients. Separate means (log MoM) were estimated for the first trimester markers in affected pregnancies at each completed week of pregnancy. In the original SURUSS report,[1] this was not done for nuchal translucency, but with the recognition that nuchal translucency medians in affected pregnancies decline with gestation from 10 or 11 weeks, SURUSS data were reanalysed to take account of this.[2] The parameters were used to estimate the risk of having a Down's Syndrome pregnancy at about 16–17 weeks, using standard methods.[3] It was not possible to determine the risk at birth because most women were screened in the second trimester, so many affected pregnancies were identified and terminated at this time. The screening performance of each test (all included maternal age) was specified as the detection rate (proportion of affected pregnancies with a positive result) and the false-positive rate (proportion of unaffected pregnancies with a positive result). This was determined using the distribution of the markers and applying the age-specific Down's Syndrome rates to the age distribution of maternities in England and Wales, from 1996 to 1998 inclusive.[3–5]

The screening performances of the main screening tests were compared. The effect of adding urine markers to these tests was examined. The detection and false-positive rates for each test change with the risk cutoff, so screening performances of the different tests are usually compared by fixing either the detection rate or the false-positive rate. The main screening tests yield high detection rates for a given false-positive rate (say 5%), so important differences in screening performance between tests can be concealed. This can be avoided by fixing the detection rate (say at 85%) and comparing the false-positive rates, which do reveal the important differences between tests.

A cost-effectiveness analysis was carried out, based on UK NHS unit costs (see Box 2)[1,6] to estimate the total cost per woman screened and per Down's Syndrome pregnancy diagnosed. The cost estimates were based on a 90% uptake rate of diagnostic tests in affected pregnancies and 80% in unaffected pregnancies (the rate is higher in affected pregnancies because such women tend to have higher risks and more often accept diagnostic testing) and based on a termination rate of Down's Syndrome pregnancies of 90%.[7] The excess risk of fetal loss from amniocentesis was taken as 0.9%.[8] Because unit costs vary, we have set up an interactive dynamic table on the website http://www.smd.qmul.ac.uk/wolfson/screencost. If local unit costs are entered, the table automatically revises and shows the new costs.

BOX 2 Unit costs per pregnancy of an NHS antenatal Down's Syndrome screening service

Screening test:	**UK cost (£)**
Quadruple	14
Combined	15
Serum Integrated	18.50
Integrated	23
Diagnostic test:	
Chorionic villus sampling with polymerase chain reaction	350
Amniocentesis without polymerase chain reaction	300
Termination of pregnancy	500
Medical evacuation of products of conception	400
Delivery	600

Results

The most effective screening test was the Integrated test, with an estimated 85% detection rate for a 0.9% false-positive rate—about one-fifth the number of false-positives that would arise using the first trimester Combined test or second trimester Quadruple test at the same detection rates (see Fig. 1). At an 85% detection rate, the false-positive rate for the Serum Integrated test was 3.9%, and 4.3% for the Combined test with the first trimester markers measured at 11 weeks. It was 6.2% for the Quadruple test, and even higher for the other tests shown in Fig. 1. If the serum Integrated test was carried out with PAPP-A measured at 10 weeks (instead of 11), the false-positive rate would be lower (2.7% instead of 3.9%). Figure 2 shows that the odds of being affected given a positive result were about five times more favourable with the Integrated test than with the Combined or Quadruple tests. Figure 3 shows the number of procedure-related unaffected fetal losses in 100,000 women screened according to specified screening tests. The Integrated test offers the safest screening and antenatal diagnosis programme, with six losses at an 85% detection rate compared, for example, with 67 with the Triple test.

The estimated screening performance of the Combined test was better than the Quadruple test and both were worse than the Integrated or Serum Integrated tests. The Double and Triple tests and nuchal translucency measurement on its own were the worst screening tests and so consideration should be given to replacing these with the better tests.

In many centres, inhibin-A is not measured in the second trimester. At an 85% detection rate, the false-positive rate for the Serum Integrated test without inhibin-A measurement was 6.0%; for the Integrated test it was 1.3%. The corresponding odds of being affected given a positive result were 1:31 and 1:7, respectively.

The median nuchal translucency, PAPP-A and free β-hCG levels in affected pregnancies change with gestational age during the first trimester, as does the standard deviation of nuchal translucency in unaffected pregnancies. These observations will affect both Down's Syndrome risk estimation and screening performance at this time in pregnancy. Table 1 shows the effect of using week-specific median nuchal translucency, PAPP-A and free β-hCG levels and week-specific standard deviations for nuchal translucency. There is a small advantage in carrying out the first trimester measurements for the Integrated test at 10 or 11 weeks rather than later, and a substantial advantage for the Serum Integrated test. The advantage of carrying out the first trimester measurements at 10 or 11 weeks rather than later has recently been reinforced with the observation that nuchal translucency MoM values in Down's Syndrome pregnancies decline with gestational age.[2] The median nuchal translucency MoMs in affected pregnancies were estimated using a log linear regression; these were 2.42, 2.18, 1.96 and 1.77 at 10, 11, 12 and 13 weeks, respectively. Although in SURUSS this trend was not statistically significant, the 10% decrease per week was almost exactly the same as the 11% decline reported by Spencer *et al.*,[9] which was highly significant ($P = 0.009$). This means that screening performance will be somewhat better at 10 and 11 weeks and somewhat worse at 13 weeks than those originally reported in SURUSS. The effect of this refinement is shown in Table 1 for the three tests which include a nuchal translucency measurement. Screening performance is better at 10 and 11 weeks, the same at 12 and worse at 13. There are few data on nuchal translucency in affected pregnancies at 10 weeks. At present, it is probably best to regard 11 weeks as the week of choice in obtaining nuchal translucency measurement in antenatal screening for Down's Syndrome. With the Combined test, the Integrated test and the Serum Integrated test, 11 weeks is therefore, on current evidence, the best time for obtaining the first trimester measurements.

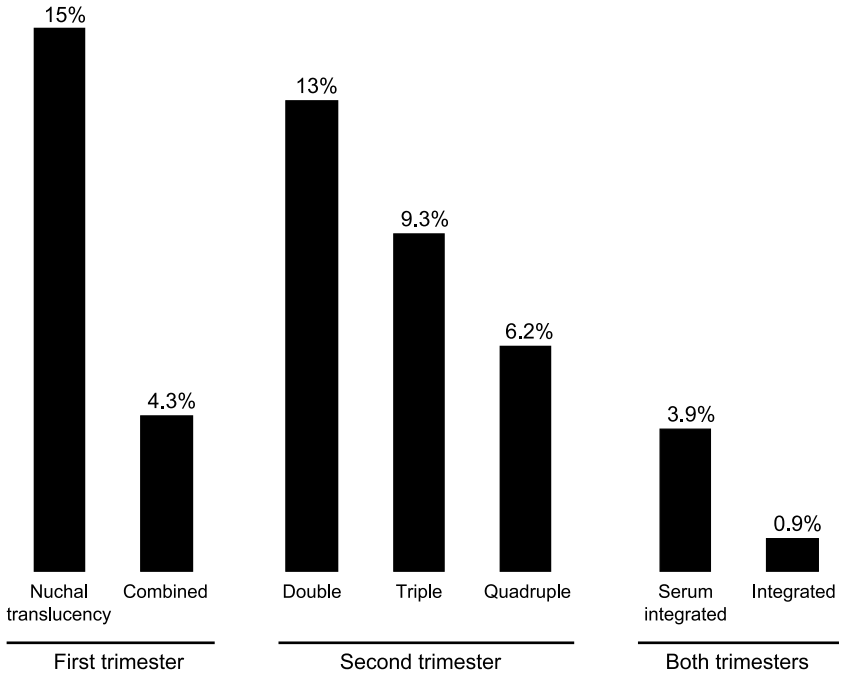

Fig. 1. False-positive rate for an 85% detection rate according to screening test in the first trimester with first trimester markers measured at 11 weeks and second trimester markers at 14–22 weeks.

Three practical results emerge in respect of nuchal translucency measurement. First, in a routine screening programme, nuchal translucency measurements are unlikely to be obtained in all pregnancies. During a 20-minute examination, a nuchal translucency measurement was not obtained in 9% of pregnancies. The success rate in obtaining a nuchal translucency measurement was greatest at 12 weeks (93%) and not much less at 10, 11 or 13 weeks. The success rate improved with sonographer experience; at 10–12 weeks from 91% in the first fifth of the study to 95% in the last fifth. This percentage is likely to be higher if nuchal translucency was used as part of a screening programme rather than for research, as sonographers would know that the measurement was to be used in the clinical management of the patient, and endeavour to spend longer on the examination if necessary. Second, there were 10

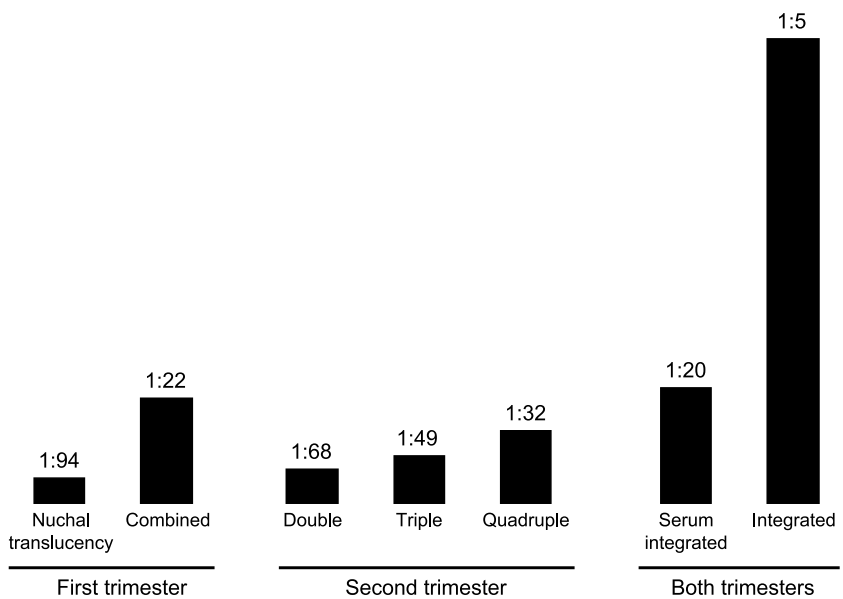

Fig. 2. The odds of being affected given a positive result for the tests shown in Fig. 1 used to achieve an 85% detection rate.

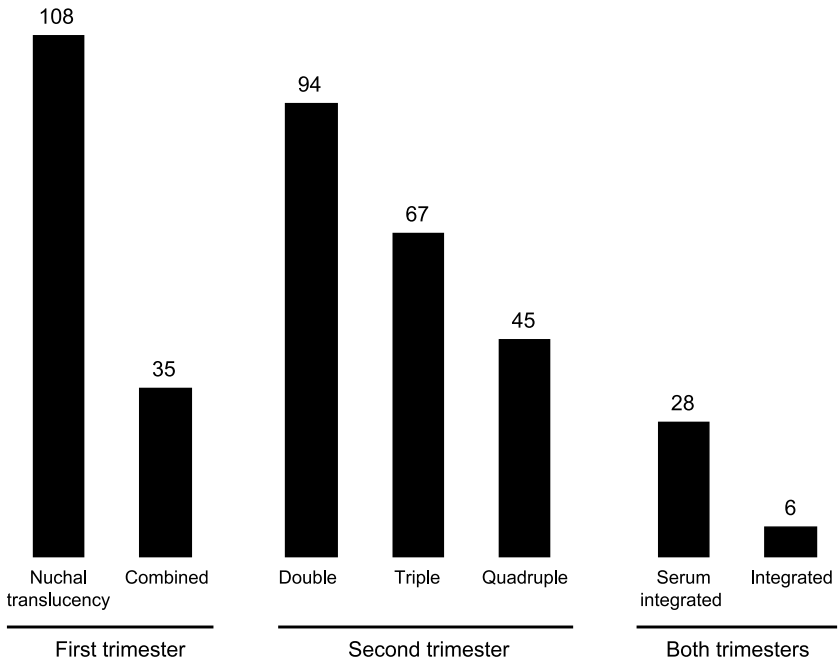

Fig. 3. The number of procedure-related unaffected fetal losses in 100,000 women screened for the tests shown in Fig. 1 used to achieve an 85% detection rate.

Down's Syndrome pregnancies at 10–13 weeks in which no nuchal translucency measurements were technically satisfactory. The median nuchal translucency in these pregnancies was 1.08 MoM, indicating little discrimination. In the 75 cases with at least one technically satisfactory measurement, the median nuchal translucency was 1.91 MoM. Without adequate quality control, affected pregnancies with poor nuchal translucency images are more likely to be missed. Third, systematic differences between sonographers meant that nuchal translucency measurements should be converted into MoMs for each sonographer rather than each centre. Using nuchal translucency alone (without maternal age), the detection rate for a 5% false-positive rate increased by 5 percentage points using sonographer rather than centre-specific data.

Table 1. The effect of first trimester gestation-specific medians on the screening performance: false-positive rate (%) for an 85% detection rate. Figures in brackets are the original SURUSS estimates that did not take account of the recently recognised decline in nuchal translucency MoMs with increasing gestation in Down's Syndrome pregnancies.

Test[a] (all with maternal age)	Gestational age (completed weeks)			
	10	11	12	13
Nuchal translucency	14 (25)	15 (22)	20 (20)	27 (20)
PAPP-A	16	26	34	42
Free β-hCG	44	35	29	23
Combined	3.5 (6.1)	4.3 (6.0)	6.0 (6.0)	7.7 (5.8)
Integrated	0.7 (1.2)	0.9 (1.2)	1.3 (1.3)	2.1 (1.5)
Serum Integrated	2.7	3.9	4.9	5.6

[a] Free β-hCG was measured rather than total hCG where either could be used.

Urine invasive trophoblast antigen was the best urinary marker and only discriminatory in the second trimester. When added to the Quadruple test it decreased the false-positive rate at an 85% detection rate from 6.2% to 4.2% and when added to the Integrated test it decreased the false-positive rate from 0.9% to 0.6%. These reductions in the false-positive rate may not be worthwhile as a separate sample has to be obtained, processed and analysed.

Figure 4 shows 'ROC curves' that give the detection rates and false-positive rates for the five most effective screening tests. The curves show the relative screening performances and the trade off between the detection rates and the corresponding false-positive rates.

Table 2 shows the estimated false-positive rate, the cost per woman screened using current NHS unit costs (including diagnosis and termination of pregnancy) and the cost per Down's Syndrome pregnancy diagnosed according to the main tests and detection rates. In the analyses, all tests include maternal age and free β-hCG is the hCG measurement used. The birth prevalence (1.74 per 1000) is based on the age distribution of maternities in England and Wales 1996–1998 resulting in an estimated second trimester prevalence of 2.26 per 1000.[4,5,10] Table 2 shows results with the first trimester markers measured at 11 weeks (and also at 10 weeks for the serum Integrated test). The average costs per woman screened with a policy set to achieve an 85% detection rate would be as follows with each of the following screening methods (all methods including age): Quadruple £29, Combined £33, Integrated £26, Serum Integrated £26. By comparison, it would be £35 for the Triple test because of the high amniocentesis rate (13%) needed to achieve an 85% detection rate. At a

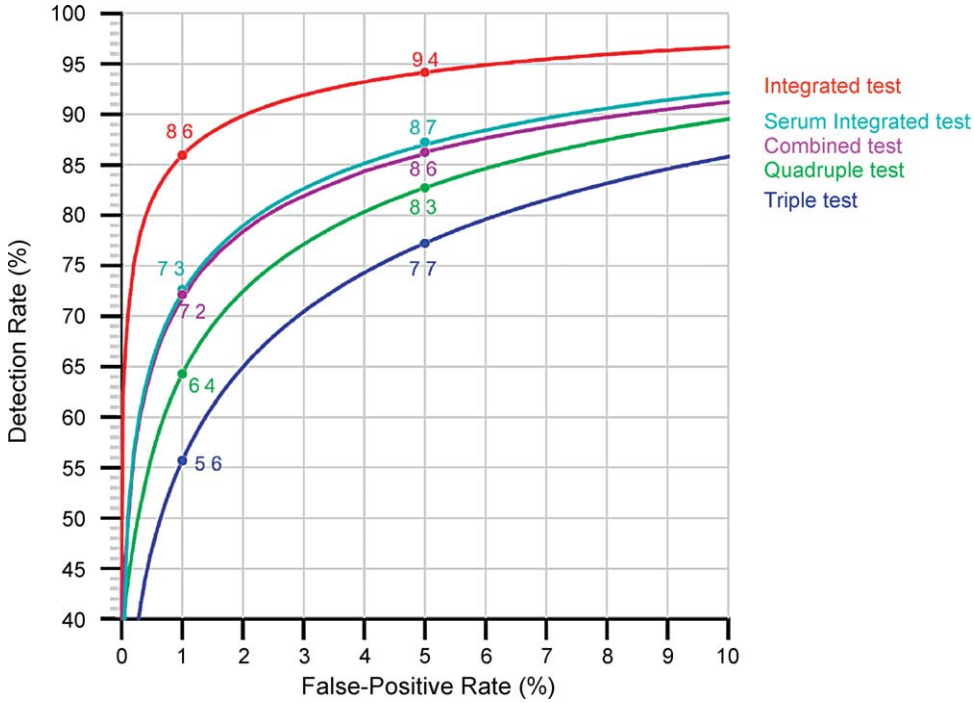

Fig. 4. Down's Syndrome detection rates and false-positive rates for specified screening tests.

Table 2. False-positive rates for specified detection rates and costs of screening (including diagnosis and termination of pregnancy) for selected screening tests according to specified detection rates using NHS unit costs. First trimester markers measured at 11 completed weeks. Figures in brackets are the estimates with PAPP-A measured at 10 completed weeks of pregnancy. The unit costs are shown in Box 2. Cost of screening based on different unit costs can be obtained from http://www.smd.qmul.ac.uk/Wolfson/screencost.

Screening test	Detection rate (%)			
	75	80	85	90
False-positive rate (%)				
Quadruple	2.5	3.9	6.2	10.6
Combined	1.3	2.3	4.3	8.4
Serum Integrated	1.3 (0.8)	2.2 (1.5)	3.9 (2.7)	7.4 (5.3)
Integrated	0.2	0.4	0.9	2.1
Cost per woman screened (£)				
Quadruple	20	23	29	40
Combined	19	22	25	39
Serum Integrated	22 (21)	24 (23)	29 (26)	37 (32)
Integrated	24	25	26	29
Cost per Down's Syndrome pregnancy diagnosed (£'000)				
Quadruple	13.2	14.4	16.8	21.7
Combined	12.6	13.6	16.1	21.5
Serum Integrated	14.5 (13.7)	15 (13.9)	16.5 (14.8)	20.2 (17.5)
Integrated	15.7	15.1	14.9	15.7

These estimates and, unless otherwise stated, others shown in this article, take account of the recently recognised decline in nuchal translucency MoMs with increasing gestation in Down's Syndrome pregnancies.

detection rate of about 85% or greater, the Integrated test and the Serum Integrated test are the most cost effective. The Integrated test, with more screening measurements, incurs more costs, but these are offset by savings achieved through the lower false-positive rate and the reduced number of amniocenteses. The rank order of cost effectiveness varies according to the detection rate.

Discussion

SURUSS provides, for the first time, a data set based on a single cohort of screened women, from which any combination of the specified screening markers within the first or second trimesters, or across both trimesters can be examined.

The main results of the SURUSS study were based on over 47,000 singleton pregnancies of which 101 had Down's Syndrome, close to what we had aimed for in the protocol (50,000 pregnancies and 100 with Down's Syndrome). SURUSS had the statistical power to yield reasonably narrow confidence intervals on the screening estimate (see tables).

The methodology used in connection with risk estimation is standard,[3,11,12] and has been empirically validated.[13,14] An alternative method is based on estimating risk for each affected and unaffected pregnancy in a study sample and counting those that exceed specified risk cutoff levels. This method is subject to substantial random error unless there are study data on much larger numbers of affected (many hundreds) and unaffected (many thousands) pregnancies than needed with the method used in SURUSS. The method adopted in SURUSS provides the most reliable estimates of screening performance.

Because of its design (avoiding screening intervention until after the second trimester serum and urine samples were collected), SURUSS avoids two sources of bias that would otherwise have affected the comparison of first and second trimester tests. In previous studies, these biases have resulted in the over-estimation of the screening performance of first trimester tests compared with tests performed in the second trimester. Both biases arise from the spontaneous fetal loss of Down's Syndrome pregnancies and both over-estimate the detection rates of earlier tests compared with later ones. The first is a general fetal loss bias that arises because Down's Syndrome pregnancies are more likely to end in a spontaneous fetal loss than unaffected pregnancies.[10] The bias can be avoided if all Down's Syndrome pregnancies (detected and missed) are ascertained after the last screening measurement to be used in any comparison. The second bias is a marker-related fetal loss bias. It arises because screening markers are associated with spontaneous fetal loss as well as with Down's Syndrome, so that detected Down's Syndrome pregnancies are more likely to end in a spontaneous fetal loss than those that are missed. In 'intervention' studies of Down's Syndrome screening, these cases are preferentially detected and result in a termination of pregnancy, and are included when, without a termination, they would have ended in a spontaneous fetal loss and would not have been included. To obtain accurate estimates of screening performance at term, this bias could only be avoided if all the screening data were available from observational studies in which no intervention was carried out on the basis of the screening markers and all the pregnancies continued to term. Such data only exist for second trimester markers based on stored antenatal serum samples before screening was introduced. SURUSS avoided this bias up to the early second trimester (as there was no material intervention until then, so comparison between first and second trimester tests were unbiased) but did not avoid it thereafter in respect of any of the markers. No study can avoid this bias to term, as a completely observational study would now be unethical. An approximate adjustment can be made to allow for the general fetal loss bias in estimating term screening performance, but this cannot be done for the marker-related fetal loss bias.

It may be helpful to illustrate how published studies on first trimester markers have been subject to fetal loss bias leading to the performance of the first trimester markers being over-estimated relative to the performance of second trimester markers and compared with performance estimates from SURUSS. The report by Bindra *et al.*[15] provides an example. It summarised screening performance using the first trimester Combined test and gave an estimate of 91.5% for the Down's Syndrome detection rate [75/82] with a 6.8% false-positive rate. The over-estimate arises because screen-detected Down's Syndrome pregnancies were ascertained at an earlier stage of pregnancy than missed cases. After making the necessary adjustments for the general fetal loss bias and setting the false-positive rate at 5%, the detection rate would be about 87% at 17 weeks of pregnancy. This is closer to the estimate of screening performance from SURUSS (83%) in the original report and the revised estimate of 86% at 11 weeks,[2] although it would be expected to be a little higher as it does not allow for the marker-related fetal loss bias between 12 and 17 weeks. Spencer *et al.*[16] reported an estimate relating to the first trimester Combined test based on 210 pregnancies with Down's Syndrome; the estimated detection rate of 89% (95% confidence interval, 85–93%) for a 5% false-positive rate, without adjustment for fetal loss bias. If such an adjustment were made, the screening performance would be less and probably statistically consistent with the SURUSS estimate. An audit of the Combined test recently reported a detection rate of 92% (23/25) for a 5.2% false-positive rate,[17] which, after the necessarily incomplete adjustment for fetal loss bias, yields a detection rate of 89% for a 5.2% false-positive rate. These comparisons with other studies show that the SURUSS estimates relating to the first trimester Combined test are in general agreement with other estimates after doing what can be done to allow for fetal loss bias. As complete allowance for such bias is not possible, the previously published estimates from first trimester intervention studies will still tend to over-estimate performance relative to the SURUSS estimate based on intervention after about 17 weeks. The difference is small but is nonetheless evident.

A meta-analysis of all existing first and second trimester screening studies would not provide a valid comparison of the screening performance of the different available tests. Apart from SURUSS, the studies are subject to the biases described above. The SURUSS results could possibly be combined with those from the US FASTER study, which has recently been completed. FASTER has a similar design to SURUSS, and so like SURUSS avoids the general fetal loss bias as well as the marker-related bias, by standardising screening performance to about 17 weeks and also standardising the estimates of screening performance to a standard age population.

Screening performance estimates at 17 weeks of pregnancy rather than at term have now become the standard. This is unavoidable because term data on all the screening tests are not available, and, for ethical reasons, are unlikely to become available because completely observational studies can no longer be conducted. This means that estimates of screening performance (for example, estimating the detection rate for a 5% false-positive rate) will be a little more favourable than previously published second trimester estimates from observational studies. For example, the Quadruple test has an estimated 83% detection rate for a 5% false-positive rate judged at about 17 weeks[1] compared with 76% judged at term.[18,19] Similarly, first trimester studies in which intervention follows the results obtained in the first trimester will yield somewhat more favourable screening results than those estimated at 17

weeks. The earlier screening performance is estimated, the better it will appear, even with tests of identical screening performance.

SURUSS confirmed that nuchal translucency measurements can be satisfactorily performed in routine obstetric practice. The mean nuchal translucency in Down's Syndrome pregnancies was 1.96 MoM, close to the estimates of 2.02 MoM from the Fetal Medicine Foundation[20] after adjustment for the general fetal loss bias. The standard deviation estimates expressed as $\log_{10}$ MoM were also similar for affected pregnancies; 0.2313 in SURUSS and 0.235 in the Fetal Medicine Foundation data. For unaffected pregnancies, it was 0.1329 in SURUSS (12–13 weeks) and 0.120 in the Fetal Medicine Foundation data. The detection rate for a 5% false-positive rate (with maternal age) was 73% in SURUSS (at 11 completed weeks[2]) and 73% in the Fetal Medicine Foundation data. de Graaf *et al.*[21] found a median nuchal translucency MoM in Down's Syndrome pregnancies of 2.03, similar to SURUSS and the Fetal Medicine Foundation. The standard deviations (0.256 and 0.152 in affected and unaffected pregnancies, respectively) were a little larger but still close. At a 5% false-positive rate, the detection rate was estimated to be 64%. The similarity of these results shows that acceptable nuchal translucency measurement can be translated from specialist to non-specialist obstetric centres and that the SURUSS estimates fit with what is already known.

As it is not possible to obtain a nuchal translucency measurement in every pregnancy within a reasonable time, it may be necessary to screen using the Serum Integrated test instead. This may arise in about 5% of pregnancies. To help avoid missed cases due to poor image quality, training, on site review of picture quality and epidemiological monitoring of nuchal translucency MoMs, standard deviations, and increase in nuchal translucency (in mm) with gestation should be instituted, in the same way that this is used to maintain biochemical assay quality. As the use of sonographer-specific nuchal translucency medians improves screening performance compared with centre-specific medians, this should be considered. The gain in performance will vary depending on the consistency in technique between sonographers, but if the workload per sonographer is sufficient (as it should be), using sonographer-specific medians is a cost-free benefit.

An observation from SURUSS is that within the first trimester nuchal translucency screening performance varied from week to week due to a decrease in the standard deviation with gestational age. As a consequence, the estimated detection rate using nuchal translucency and maternal age increased with gestation. The median MoM value in Down's Syndrome pregnancies also tended to decline with gestation (by 10% per week) but this was not statistically significant. After the publication of the SURUSS report, Spencer *et al.*[22] found an 11% decline per week, which was statistically significant. With the recognition that nuchal translucency MoMs in Down's Syndrome pregnancies decline, it has become apparent that screening performance of nuchal translucency measurement at 10–11 weeks was greater than at 12 and 13,[2] the effect of the decline in nuchal translucency MoMs in Down's Syndrome pregnancies outweighs the reduction in the standard deviation in unaffected pregnancies. Table 1 shows the revised estimates of the false-positive rates for an 85% detection rate[2] taking account of both the declining nuchal translucency MoM and nuchal translucency standard deviation with increasing gestational age. The original SURUSS report estimates[1] are shown in brackets.

The median PAPP-A in affected pregnancies becomes less discriminatory (that is, closer to 1.0 MoM) as the gestational age advances, while the median in free β-hCG MoM becomes more discriminatory.

There were 10 affected pregnancies with PAPP-A measurements at 10 weeks in SURUSS. We present screening performance for the Serum Integrated test at 10 weeks as there is sufficient information in the published literature that shows that, on average, PAPP-A levels in affected pregnancies are lower at 10 weeks than at 11. The regression used to estimate the normal medians in SURUSS is therefore reasonably secure at 10 weeks as well as at 11 weeks.

As a result of these changes in the screening markers, the first trimester cannot be treated as a single time period in the same way as the second trimester. A woman's risk and the associated screening performance of the test will depend on when she is screened during the 10–13 week period. The optimal time is 10–11 weeks, but because of the sparse nuchal translucency data at 10 weeks in affected pregnancies, we give estimates in this paper for 11 completed weeks.

Before implementing screening programmes using a particular screening test, a risk cutoff needs to be selected. Table 3 shows the estimated screening performance according to risk cutoff and according to test. Different tests require different risk cutoffs to achieve a high screening performance so a universal cutoff is not appropriate. Once a cutoff is selected for a particular test, the detection and false-positive rates are fixed. So, for example, it is not possible to specify a 1 in 250 risk cutoff for, say, the Quadruple test and also specify a false-positive rate of 5% or less. If the latter was specified as a standard, a cutoff of 1 in 200 would have to be selected. It is also important to recognise that the age distribution of the population is a factor that influences the detection rate and false-positive rate at a given cutoff. Among older populations, both the detection rate and the false-positive rate will be higher at a given risk cutoff.

With the main screening tests available, improving the odds of being affected given a positive result is mainly achieved by reducing the false-positive rate associated with the more effective tests (such as the Integrated test, which loses little in detection in spite of the lower false-positive rate). This can be achieved by raising the risk cutoff (to, say, 1 in 100 compared with 1 in 300).

Table 3. Screening performance according to second trimester risk cutoff and screening test.

Screening test (all include maternal age)	Risk cutoff														
	1 in 100			1 in 150			1 in 200			1 in 250			1 in 300		
	DR (%)	FPR (%)	OAPR	DR (%)	FPR (%)	OAPR	DR (%)	FPR (%)	OAPR	DR (%)	FPR (%)	OAPR	DR (%)	FPR (%)	OAPR
Integrated[a]	87	1.2	1:6	89	1.6	1:8	90	2.1	1:10	91	2.6	1:12	92	3.0	1:14
Serum Integrated[a]															
PAPP-A at 10 weeks*	82	1.8	1:10	85	2.6	1:14	87	3.3	1:17	88	4.0	1:20	89	4.6	1:23
PAPP-A at 11 weeks	79	2.0	1:11	82	2.9	1:16	85	3.8	1:20	86	4.6	1:23	87	5.3	1:27
Combined[a]	77	1.8	1:10	81	2.6	1:14	83	3.4	1:18	85	4.2	1:22	86	4.9	1:25
Quadruple	74	2.5	1:15	79	3.6	1:20	82	4.7	1:25	84	5.7	1:30	86	6.6	1:34

DR = detection rate; FPR = false-positive rate; OAPR = odds of being affected given a positive result.
Free β-hCG, rather than total hCG measured in tests that use second trimester markers.
[a] All with first trimester markers at 11 weeks except *.

With the availability of both first and second trimester screening, a stepwise approach offering all women the Combined test and then in those who are screen-negative, the Quadruple test, may seem to be an effective policy. Table 4 shows that such a policy would lead to a detection rate of 94% and a false-positive rate of 9.0%. The Integrated test with a risk cutoff level to achieve the same 94% detection rate yields a much lower false-positive rate (4.9%) with a consequent significant improvement in the safety and substantial cost savings. Stepwise screening should therefore be avoided.

Conclusion

Our results show that, on the basis of efficacy, safety and cost, the Integrated test is the test of choice, confirming previous estimates based on combining the results from different studies.[23] Adding other markers provided little benefit. If a nuchal translucency measurement was unavailable, the Serum Integrated test would be the next best screening method, with a better screening performance than any first or second trimester serum screening test. A multicentre demonstration project in which about 25,000 women received integrated screening has shown that the Integrated test was acceptable in public and private screening programmes using various forms of Integrated tests (for example, with and without nuchal translucency or with the Triple test in the second trimester instead of the Quadruple test) (N Wald, unpublished observation). Using the polymerase chain reaction or fluorescent *in situ* hybridisation, a diagnosis can be made at 15 weeks.

For women who present for the first time in the second trimester of pregnancy, the Quadruple test is the test of choice, as previously documented in a systematic review[7] and shown in an audit of a screening programme using this test.[24] For women who request a screening result and a diagnosis made before 14 weeks of pregnancy, the Combined test is the best option, although at the cost of a loss in detection and a higher false-positive rate than the Integrated test. A summary of the screening recommendations is shown in Table 5.

The transition from current screening practice to recommended practice is achievable. Currently, most screening

Table 4. Screening using a stepwise or integrated approach.

Policy based on	Detection rate[a]	False-positive rate[a]	In 100,000 women screened		
			No. of false-positives	No. of unaffected fetal losses	Cost of screening
Stepwise screening[b]	94%	9.0%	9026	81	£5.2 million
Integrated screening[c]	94%	4.9%	4898	44	£3.6 million
Advantage of Integrated screening		46% lower	4128 fewer false-positives	37 fewer unaffected fetal losses	£1.6 million less expensive

[a] Assumes 100% uptake of chorionic villus sampling or amniocentesis after each test result.
[b] Women have first trimester Combined test and screen negatives have second trimester Quadruple test (risk cutoff 1 in 250 for both tests).
[c] All women have first trimester nuchal translucency, PAPP-A and second trimester Quadruple test (risk cutoff set to achieve same detection rate as stepwise policy).

Table 5. Efficacy and safety of the recommended screening tests (first trimester markers measured at 11 completed weeks).

Test		Efficacy		Safety
		FPR (%) for an 85% DR	OAPR	No. of procedure-related fetal losses per 100,000
Offered routinely				
In general:	Integrated	0.9	1:5	6
If NT not available*	Serum Integrated	3.9 (2.7)	1:20 (1:14)	28 (19)
If first visit is in second trimester	Quadruple	6.2	1:32	45
Available but not offered routinely				
For women who request a first trimester test	Combined	4.3	1:22	35

FPR = false-positive rate; DR = detection rate.
* Numbers in brackets related to PAPP-A measured at 10 weeks instead of 11.

centres will be offering serum screening. Throughout the world this screening is mainly carried out using the Triple test. Adding inhibin-A (to form the Quadruple test) is simple, and many laboratories have done so. Indeed, we have successfully screened about 50,000 pregnancies over the past seven years with the Quadruple test.[24] The PAPP-A assay is a simple analytical addition (to form the Serum Integrated test), but using this test involves collecting blood at about 10–11 weeks as well as at about 15 weeks. This may require reorganising antenatal care arrangements, but the health gain resulting from this is significant and it is worth making the change. Providing an ultrasound service performing reliable nuchal translucency measurements has been achieved in many centres and in many others it could be installed, but in some rural areas the Serum Integrated test may have to be the standard.

With the Integrated test, one affected pregnancy will be found in about every 6 women who have an amniocentesis instead of one in about every 50 as is the case with the Triple test (see Fig. 2). Also, the risk estimates in affected pregnancies will tend to be high (about one in three on average with the Integrated test and one in four with the Serum Integrated test), while the risk estimates in unaffected pregnancies will tend to be low (about 1 in 24,000 on average with the Integrated test and 1 in 13,000 with the Serum Integrated). Both effects reflect the wider separation in risk estimates between affected and unaffected pregnancies with higher performance screening tests.

The benefits of the recommended tests arise because of the higher screening detection rate, fewer false-positives and hence greater safety because of the need for fewer invasive diagnostic procedures. With an appropriate screening policy and reasonable cost control, this gain in quality of care can be achieved cost effectively, that is, without an increase in the cost of detecting an affected pregnancy compared with using existing screening methods and without an increase in the cost per woman screened at a fixed detection rate.

Screening which can provide high risk estimates in affected pregnancies means that more women with such pregnancies who are screen-positive will choose to have a diagnostic amniocentesis. Several studies have shown that, among women with screen positive results, the uptake of amniocentesis is related to the estimated risk of having a Down's Syndrome pregnancy. For example, in a screening programme based on the Triple test, the uptake was 72% if the risk was >1:10 and 48% between 1:135 and 1:270.[25]

In summary, therefore, the SURUSS results show that in antenatal screening for Down's Syndrome, it is now possible to detect about 9 out of every 10 affected pregnancies with a false-positive rate of 1–2%, substantially lower than in the past, so achieving a significantly higher level of safety by reducing the number of women offered an invasive diagnostic test such as amniocentesis.

In 2001 in an Editorial in this Journal entitled 'Down's Syndrome screening: where to now?' Whittle[26] considered the variety of screening methods and the confusion this causes and expressed justified concern over our failure to recommend the most effective and safe methods. In the same year, Yvette Cooper,[27] UK Minister of Public Health (and six months pregnant at the time) said in the Guardian newspaper 'Different units offer different tests at different points in pregnancy. In fact if you switch hospitals at the right point, you can probably clock up six different tests for Down's Syndrome (scans, double and triple tests galore) and each will give you a slightly different result)'. Whittle concludes his editorial saying 'We must not fail again'. The necessary information is now at hand so that the appropriate screening recommendations can be made for health care providers such as the NHS, so that women can be offered the most effective and safe methods of antenatal screening for Down's Syndrome.

Acknowledgements

The SURUSS project was supported under the UK Health Technology Assessment research and development programme.

Declaration of interests

St Bartholomew's and the Royal London School of Medicine and Dentistry together with the Foundation for Blood Research and Women and Infants Hospital (Providence, Rhode Island) hold a patent for unconjugated oestriol measurement as a marker for Down's Syndrome screening. Professor Wald is a director of Logical Medical Systems, which produces αlpha, a commercial interpretive software package for Down's Syndrome screening using ultrasound and serum markers. He is also a director of Intema, which holds patent interests (granted and pending) for the Integrated test.

N.J. Wald,[a] C. Rodeck,[b] A.K. Hackshaw,[a] A. Rudnicka[a]
[a]*Department of Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, UK*
[b]*Department of Obstetrics and Gynaecology, University College London, UK*

References

1. Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *J Med Screen* 2003;**10**:56–104. Available: http://www.ingentaselect.com/rsm/09691413/v10n2/contp1-1.htm.
2. Wald N, Rodeck C, Rudnicka A, Hackshaw A. Nuchal translucency and gestational age. *Prenat Diagn* 2004;**24**:150–153.
3. Wald N, Hackshaw A. Tests using multiple markers. In: Wald NJ, Leck I, editors. *Antenatal and Neonatal Screening*, *2nd edition*. Oxford: Oxford University Press, 2000:23–57.
4. Morris JK, Mutton DE, Alberman E. Revised estimates of maternal age specific live birth prevalence of Down's syndrome in the absence of antenatal screening and selective termination. *J Med Screen* 2002;**9**:2–6.
5. Office for National Statistics. *Birth Statistics, Series FM1, Numbers 25–27*. London: HMSO, 1996, 1997, 1998.
6. Gilbert RE, Augood C, Gupta R, et al. Screening for Down's syndrome: effects, safety, and cost effectiveness of first and second trimester strategies. *BMJ* 2001;**323**:423–425.
7. Wald NJ, Kennard A, Hackshaw A, McGuire A. Antenatal screening for Down's syndrome. *J Med Screen* 1997;**4**:181–247.
8. Gosden C, Tabor A, Leck IL et al. Amniocentesis and chorionic villus sampling. In: Wald N, Leck I, editors. *Antenatal and Neonatal Screening*, *2nd edition*. Oxford: Oxford University Press, 2000:470–516.
9. Spencer K, Bindra R, Nix ABJ, Heath V, Nicolaides KH. Delta-NT or NT MoM: which is the most appropriate method for calculating accurate patient-specific risks for trisomy 21 in the first trimester? *Ultrasound Obstet Gynaecol* 2003;**22**:142–148.
10. Morris JK, Wald NJ, Watt HC. Fetal loss in Down syndrome pregnancies. *Prenat Diagn* 1999;**19**:142–145.
11. Wald NJ, Cuckle HS, Densem JW, et al. Maternal serum screening for Down's syndrome in early pregnancy. *BMJ* 1988;**297**:883–887.
12. Royston P, Thompson SG. Model-based screening by risk with application to Down's syndrome. *Stat Med* 1992;**11**:257–268.
13. Wald NJ, Huttly WJ. Validation of risk estimation using the quadruple test in prenatal screening for Down syndrome. *Prenat Diagn* 1999; **19**:1083.
14. Wald NJ, Hackshaw AK, Huttly W, Kennard A. Empirical validation of risk screening for Down's syndrome. *J Med Screen* **3**:185–187.
15. Bindra R, Heath V, Liao A, Spencer K, Nicolaides KH. One-stop clinic for assessment of risk for trisomy 21 at 11–14 weeks: a prospective study of 15 030 pregnancies. *Ultrasound Obstet Gynecol* 2002;**20**:219–225.
16. Spencer K, Souter V, Tul N, Snijders R, Nicolaides KH. A screening programme for trisomy 21 at 10–14 weeks using fetal NT, maternal serum free β-hCG and PAPP-A. *Ultrasound Obstet Gynecol* 1999;**13**: 231–237.
17. Spencer K, Spencer CE, Power M, Dawson C, Nicolaides KH. Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemical in a one-step clinic: a review of three years prospective experience. *Br J Obstet Gynaecol* 2003;**110**:281–286.
18. Wald NJ, Densem JW, George L, Muttukrishna S, Knight PG. Prenatal screening for Down's syndrome using inhibin-A as a serum marker. *Prenat Diagn* 1996;**16**:143–153.
19. Wald NJ, Densem JW, George L, et al. Inhibin-A in Down's syndrome pregnancies: revised estimate of standard deviation. *Prenat Diagn* 1997;**17**:285–290.
20. Nicolaides KH, Snijders RJM, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening. *Prenat Diagn* 1998;**18**:519–521.
21. de Graaf IM, Pajkrt E, Bilardo CM, Leschot NJ, Cuckle HS, van Lith JMM. Early pregnancy screening for fetal aneuploidy with serum markers and nuchal translucency. *Prenat Diagn* 1999;**19**:458–462.
22. Spencer K, Bindra R, Nix AB, Heath V, Nicolaides KH. Delta-NT or NT MoM: which is the most appropriate method for calculating accurate patient-specific risks for trisomy 21 in the first trimester? *Ultrasound Obstet Gynecol* 2003;**22**:144–148.
23. Wald NJ, Watt HC, Hackshaw AK. Integrated screening for Down's syndrome based on tests performed during the first and second trimesters. *N Engl J Med* 1999;**341**:461–467.
24. Wald NJ, Huttly WJ, Hackshaw AK. Antenatal screening for Down's syndrome with the quadruple test. *Lancet* 2003;**361**:835–836.
25. Chen J, Heffley D, Beazoglu T, Benn PA. Utilization of amniocentesis by women screening positive for Down syndrome on the second-trimester Triple test. *Community Genet* 2000;**3**:24–30.
26. Whittle M. Down's syndrome screening: where to now? *Br J Obstet Gynaecol* 2001;**108**:559–561.
27. Cooper Y. In favour of motherhood. *Guardian* 2001 (April 30).