An update on antenatal screening for Down’s syndrome and specific implications for assisted reproduction pregnancies

Boaz Weisz1,3 and Charles H.Rodeck2

1Department of Obstetrics and Gynaecology, Sheba Medical Center, Tel-Hashomer, Israel and 2Department of Obstetrics and Gynaecology, University College London, London, UK

3To whom correspondence should be addressed at: Department of Obstetrics and Gynaecology, Sheba Medical Center, Tel-Hashomer 52321, Israel. E-mail: boazmd@zahav.net.il

Since the introduction of antenatal serum screening for Down’s syndrome (DS) more than two decades ago, several screening approaches have been utilized in routine clinical practice. The current DS screening strategies involve mid-trimester serum biochemistry tests, first trimester tests combining sonographic markers and serum biochemistry and integration of first and second trimester markers. In this review, we evaluate the performance of DS screening strategies according to the Serum, Urine and Ultrasound Screening Study (SURUSS), the First and Second Trimester Evaluation of Risks (FASTER) Trial and the Serum Biochemistry and Fetal Nuchal Translucency Screening (BUN) Study. We also evaluate the performance of first trimester screening in studies and meta-analyses by other groups. Specific issues related to assisted reproduction technology (ART) pregnancies are also addressed in this review.

Key words: Down’s syndrome/integrated screening/nuchal translucency/screening

Introduction

Since the introduction of antenatal serum screening for Down’s syndrome (DS) more than two decades ago (Cuckle et al., 1984), several screening approaches have been utilized in routine clinical practice. The original method for screening women of advanced maternal age involved invasive testing that was offered to 5% of the population and identified fewer than 30% of fetuses with DS. Subsequently, newer tests incorporating ultrasound and/or maternal serum biochemistry have been introduced to improve the detection rate (DR—proportion of affected pregnancies with positive results) and reduce the number of unnecessary invasive procedures. The current DS screening strategies involve the more traditional second trimester serum biochemistry tests, first trimester tests that combine sonographic markers and serum biochemistry and integration of first and second trimester markers.

In this review, we will discuss the concept of DS screening, evaluate the performance of DS screening strategies and present specific issues with relevance to assisted reproduction pregnancies.

Concept of screening

Screening is a systematic application of a test or enquiry to identify subjects at sufficient risk of a specific disorder so that they can benefit from further investigation or direct preventive action (Wald, 1994). There is no universally accepted definition of medical screening, but there is a general agreement that it should contain three elements: (i) It should identify those individuals who are at sufficient high risk of a specific disorder, preliminary to a diagnostic test and, if required, preventive action. (ii) It should be offered systematically to people with no signs or symptoms of the disease for which screening is being conducted. (iii) It should be beneficial to the individuals being screened. Routine screening in modern reproductive medicine was initiated as serum screening for open neural tube defects (NTD) (Wald et al., 1977). This, followed by preventive measures (folic acid uptake), has dramatically reduced the prevalence of NTD in England and Wales (Wald and Leck, 2000). While screening for NTD has remained relatively unaltered during the last three decades, screening for DS has evolved substantially and is still a matter of debate in national clinical policies.

The three elements of screening apply also to DS where screening is offered systematically to every pregnant woman and selection is for high-risk pregnancies, which are then offered an invasive diagnostic test, that is, chorionic villus sampling (CVS) or amniocentesis. Beneficence relates to the ability to choose whether to discontinue an affected pregnancy and the ability to prepare for a delivery of a DS baby for those who choose to continue the pregnancy.

Assessing the abundant screening tests available today is done by evaluating the DR of each test, the false-positive rate (FPR—proportion of unaffected pregnancies with positive results) and the odds of being affected, given a positive result (OAPR—the ratio of the number of affected to unaffected individuals with positive results, i.e. affected positive : unaffected positive). The OAPR is equivalent to the positive predictive value (PPV), but it conveys a
clearer impression of the performance of the test when the PPV is high. A better test has a higher DR and OAPR and lower FPR.

Comparisons between tests cannot be done by evaluating a single factor such as the DR or the FPR. For every test, increasing the DR will increase, unavoidably, the FPR. Therefore, to compare DR between tests, the FPR should be controlled for. A common presentation used for comparison of the DR between several tests is the DR 5, which is the DR for 5% FPR. Alternatively, in cases where low FPR is important due to an inherent risk in the diagnostic test (i.e. possible pregnancy loss), the FPR is compared for a constant DR. For example, FPR 85 and FPR 90 represent the FPRs for 85 and 90% DR, respectively. Comparison of DS screening between tests performed in the late first or early second trimesters should take into account the bias of spontaneous miscarriage of DS pregnancies and the bias of markers predicting miscarriage or loss in chromosomally normal pregnancies.

Many studies have evaluated the performance of each screening modality. The literature is abundant with papers describing the performance of each screening test. However, there is a paucity of information that clearly compares first and second trimester tests and that can be used for guidance for physicians and patients to choose the most appropriate test. The two major multi-centre studies that have compared several first and second trimester screening modalities are the Serum, Urine and Ultrasound Screening Study (SURUSS) (Wald et al., 2003b, 2004) and the First and Second Trimester Evaluation of Risks (FASTER) Trial (Malone et al., 2005a). The SURUSS, which aimed to determine the most effective, safe and cost-effective method of antenatal screening for DS, involved 25 maternity units, and results were based on 47 053 singleton pregnancies, including 101 DS patients. The FASTER Trial was conducted in 15 centres throughout USA, recruiting 42 367 women, and results were based on 38 033 eligible singleton pregnancies with 117 DS patients. The comparison of the various screening tests in this review is based on the SURUSS and FASTER Trial, as well as on other studies and meta-analyses reporting the performance of first trimester screening (Wapner et al., 2003; Nicolaides et al., 2005).

Second trimester screening tests

Screening for DS by maternal age started three decades ago when aneuploidy was offered only to older women. Finding an association between elevated serum α-fetoprotein (AFP) and fetal chromosomal abnormalities, including DS (Merkatz et al., 1984), led to a single-marker screening test using AFP along with maternal age (Cuckle et al., 1984, 1987). Subsequently, the association between elevated serum hCG and DS led to the double test (AFP and hCG). The third marker for DS was unconjugated estriol (uE 3), which was found to be lower in affected pregnancies (Canick et al., 1984). This led to the triple test, which is still very commonly used (Wald et al., 1987). The quadruple test, the combination of AFP, hCG, uE 3 and Inhibin, is currently the most popular second trimester screening test in the USA (D’Alton and Cleary-Goldman, 2005). The DR 5 and the FPR 85 for each of the second trimester tests are presented in Table I.

<table>
<thead>
<tr>
<th>Test</th>
<th>DR 5 (%)</th>
<th>FPR 85 (%)</th>
<th>OAPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double test</td>
<td>71</td>
<td>13</td>
<td>1 : 68</td>
</tr>
<tr>
<td>Triple test—SURUSS</td>
<td>77</td>
<td>9</td>
<td>1 : 49</td>
</tr>
<tr>
<td>Triple test—FASTER</td>
<td>69</td>
<td>14</td>
<td>NA</td>
</tr>
<tr>
<td>Quadruple test—SURUSS</td>
<td>83</td>
<td>6</td>
<td>1 : 32</td>
</tr>
<tr>
<td>Quadruple test—FASTER</td>
<td>81</td>
<td>7</td>
<td>1 : 37</td>
</tr>
</tbody>
</table>

DR 5, detection rate for 5% FPR. FASTER, First and Second Trimester Evaluation of Risks Trial; FPR 85, false-positive rate for 85% DR; NA, not applicable; OAPR, the ratio of the number of affected to unaffected individuals with positive results; SURUSS, Serum, Urine and Ultrasound Screening Study.

Interpretation of the results of the second trimester screening tests by the SURUSS shows that the quadruple test is superior to the triple test by increasing the DR 5 by only 1.07-fold. However, a more important observation is that for a similar DR, the need of invasive tests is reduced by 35%. Assuming that each woman with a positive screening test chooses to have an amniocentesis, women tested by the quadruple test have a 35% lower risk of requiring amniocentesis relative to the triple test. Many centres quote 1% excess risk of pregnancy loss after amniocentesis (Tabor et al., 1986). However, recent (non-randomized) studies showed a relatively smaller risk that varies between 0.6% (Seeds, 2004) and 0.2% (Nassar et al., 2004). With these more recent figures, changing the policy of DS screening from the triple test to the quadruple test would result in 6–18 less fetal losses per 100 000 pregnancies. The FASTER Trial shows a much greater difference in DR and FPR between the triple and quadruple tests.

First trimester screening

First trimester screening for DS is a relatively novel practice. The major breakthrough of early screening was the identification (Szabo and Gellen, 1990; Nicolaides et al., 1992) and implementation (Pandya et al., 1995) of nuchal thickness (NT) measurement at 11–14 weeks’ gestation. The performance of NT screening varies between studies. In the SURUSS, the DR 5 of the NT (and maternal age) at 12 weeks was 69% and the FPR 85 was 20% (Wald et al., 2003b). Similar results were published in other two multi-centre studies. In the Serum Biochemistry and Fetal Nuchal Translucency Screening Study (BUN Study), the DR 5 was 69% and the FPR 85 was about 15% (Wapner et al., 2003). In the FASTER Trial, the DR 5 was 68% and FPR 85 was 23% (at 12 weeks). The conclusion of these studies was that although NT has comparable DR to second trimester tests, it is accompanied by a relatively high FPR. However, other studies presented better results, and a review of 19 prospective studies (including about 200 000 patients) suggested a higher DR of 77% for 4% FPR (Nicolaides, 2004). A recent single-centre prospective study that included 30 564 pregnancies presented a DR 5 of 82% with a FPR 85 of 8% (Avgidou et al., 2005). The differences between the studies are not entirely understood but might in part reflect more meticulous NT measurements in single-centre studies compared to national and multi-centre studies.

During the early 1990s, several studies reported the association between DS and low levels of pregnancy-associated plasma
Antenatal screening for Down’s syndrome

protein A (PAPP-A) (Wald et al., 1992) and high levels of hCG (Spencer et al., 1992) and the use of these for screening in the first trimester (Forest et al., 1997). The combinations of NT measurement (11+0 to 13+6 weeks) with these two serum biochemical markers in the first trimester compose the combined test (Wald and Hackshaw, 1997). The incorporation of these two serum markers to the NT measurement caused a significant and important decrease in the FPR of first trimester screening. By combining these two serum markers with the NT, the FPR of the combined test is the availability of the results in the late first trimester, enabling karyotyping by CVS and early surgical termination of pregnancy, when indicated.

Several sonographic markers in the first trimester might increase the sensitivity and specificity of early screening tests. Fetuses with DS have an increased resistance (higher PIV—pulsatility index for veins and negative a-wave) in the ductus venosus (DV) during early pregnancy. It is estimated that the addition of DV Doppler studies would increase the DR of the NT measurement by 11% and of the combined test by 4% (yielding a DR of up to 92% by combined test and DV) (Borrell et al., 2005). However, such a test requires specific skills and is time consuming; therefore, it is not considered as a screening test for the general population. Another promising marker of DS is the absence of nasal bone (NB) on the first trimester scan (Cicero et al., 2001). It was estimated that combining this marker with NT measurement and serum markers (NB+NT+PAPP-A+β-hCG) might increase the DR up to 97% (Cicero et al., 2003). However, other studies challenged the reproducibility of this test (Senat et al., 2003). The FASTER study found poor correlation between the absence of NB and DS and concluded that first trimester NB evaluation was not a useful test for population screening for trisomy 21 and that it added little to first trimester NT screening (Malone et al., 2004). The difficulty in performing first trimester NB sonography consistently, in the general population setting, will significantly limit the usefulness of this screening technique. It is possible that due to specific teaching and guidance of NB scanning in the first trimester, this technique will be reserved for high-risk cases.

First and second trimester screening

The integration of first and second trimester screening markers in order to report one risk factor was initiated in the late 1990s (Wald et al., 1999a). The assumption was to use each marker at its optimal time point, that is, when the difference between the DS cases and controls is largest. The integrated test combines first trimester NT measurement and serum PAPP-A levels with second trimester AFP, β-hCG, E_{3} and Inhibin (quadruple test). In the initial report (Wald et al., 1999a), the estimated DR of the integrated test was 94%. This is similar to the DR results of the SURUSS and FASTER Trial of 93 and 95%, respectively (Wald et al., 2003a; Malone et al., 2005a). The major advantage of this test is the low OAPR (1 : 6 in SURUSS), which implies that with the integrated test fewer women will need to undergo invasive testing with its inherent risk of miscarriage, and equally importantly, fewer women are made anxious about their pregnancy (Marteau et al., 1993). The integrated test has been challenged ethically since the integration of first and second trimester markers in a single test could pose the problem of withholding first trimester results and thus denying the possible advantages of an earlier pregnancy termination (Canini et al., 2002). A recent study has shown that compared to pregnant women, health-care professionals place a higher value on earlier tests. This concern may result in screening policies that favour timing in the selection of a test and neglect tests, associated with lower miscarriage rates and higher DRs, conducted later in pregnancy (Bishop et al., 2004). Indeed, a study of pregnant women who underwent the integrated test (after having a second trimester test in their previous pregnancy) concluded that women receiving prenatal care are prepared to wait until the second trimester for more accurate DS risk estimates on which to base their decision-making (Palomaki et al., 2005). This issue was addressed by the FASTER Trial, which suggested another policy of ‘stepwise sequential screening’: women with positive first trimester combined test are offered CVS, while those with low-risk results continue to have the quadruple test at 15 weeks’ gestation, and a new risk estimate is provided by combining the first and second trimester variables. This approach yields a DR of 95% for 5% FPR (while the integrated test yields the same DR for 4% FPR). This represents a small trade-off in the FPR for earlier interventions in high-risk pregnancies. Another issue that was raised is that in national screening programmes, some women will fail to attend for the second blood analysis. The integrated test has not yet been implicated in routine large-scale screening programmes. However, an initial report of screening by the integrated test in a teaching hospital in central London is promising (C.H.Rodeck, personal communication), and the rate of not attending for the second blood analysis is about 5%. This rate is even lower in hospitals that have a routine antenatal visit at 16 weeks of gestation.

In cases when the NT is not measured, either due to technical difficulties or due to lack of expert sonographers, it is possible to perform the serum integrated test (PAPP-A in the first trimester and quadruple test in the second trimester), which also yields a relatively high DR (Knight et al., 2005).

<table>
<thead>
<tr>
<th>Study</th>
<th>DS/eligible patients</th>
<th>DR_{5} (%)</th>
<th>FPR_{85} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURUSS</td>
<td>101/47 053</td>
<td>83</td>
<td>6</td>
</tr>
<tr>
<td>BUN</td>
<td>61/8 514</td>
<td>79</td>
<td>9</td>
</tr>
<tr>
<td>FASTER</td>
<td>92/36 120</td>
<td>85*</td>
<td>5*</td>
</tr>
</tbody>
</table>

BUN, Serum Biochemistry and Fetal Nuchal Translucency Screening Study; DS, Down’s syndrome; DR_{5}, detection rate for 5% FPR; FASTER, First and Second Trimester Evaluation of Risks Trial; FPR_{85}, false-positive rate for 85% DR; SURUSS, Serum, Urine and Ultrasound Screening Study.
*At 12 weeks.
B. Weisz and C. H. Rodeck

An important aspect of the integrated test is that it links with the gold-standard diagnostic test (i.e. amniocentesis). With rapid diagnostic testing available nowadays (i.e. PCR or fluorescence in-situ hybridization [FISH]), final diagnosis is made by 16 weeks rather than 19 or 20 weeks and well before fetal movements are felt by the pregnant woman.

Current screening programs

In the UK, the National Screening Committee evaluated the DS screening service in 1998 and found deficiencies such as inequity in provision of services, variation in type and quality of service, an absence of performance standards and a lack of capacity to support a nationwide programme (National Down’s Syndrome Screening Programme for England, 2004). In order to address these issues, in October 2003 the National Institute for Clinical Excellence (NICE) recommended that screening programs should have DR of at least 60% by 2004-05 and that by April 2007 a DR of greater than 75% with a FPR of less than 3% should be achieved (National Collaborating Centre for Women’s and Children’s Health—Commissioned by the National Institute for Clinical Excellence [NICE, 2005]). At present, the available tests that meet the criteria for 2007 are the combined test, the Integrated or the serum integrated test (if nuchal translucency measurement is unavailable) and the quadruple test for women who attend for screening after the first trimester.

Current policies in the USA vary, and there are no national guidelines. A survey conducted in 2001 reported that the most common method of screening was the triple test (Egan et al., 2002) and more recently the quadruple test (D’Alton and Cleary-Goldman, 2005). It is reasonable that the results of the SURUSS, BUN and FASTER studies may change the trend in favour of first trimester or an integration of first and second trimester screening. Indeed, in 2004, the American College of Obstetrics and Obstetricians (ACOG Committee Opinion #296, 2004) issued a Committee Opinion stating that first trimester screening may be considered provided that there is (i) appropriate training and monitoring programs, (ii) sufficient counselling to women regarding the differences between the tests and (iii) access to an appropriate diagnostic test (i.e. CVS) when screening test results are positive.

Specific issues for assisted reproduction technology (ART) pregnancies

Pregnancies conceived by ART carry a higher psychological and financial burden compared to spontaneous pregnancies (Oddens et al., 1999). The uptake of amniocentesis in ART pregnancies is believed to be lower compared to controls, mainly due to the inherent risks of amniocentesis in these ‘more precious’ pregnancies (Geipel et al., 2004). Therefore, this population gains specifically from the low OAPR of the integrated test, which reduces their risk of being high risk for trisomy 21 and having an invasive test. Other issues, which apply to ART pregnancies, are (i) a relatively older population of pregnant women, (ii) differences in serum markers compared to spontaneous pregnancies and (iii) a higher rate of multi-fetal pregnancies.

Maternal age in ART pregnancies

The proportion of women aged 35 years or more in some IVF clinics is approximately 50% (Gissler et al., 2004). Inherently, an older population of women would gain from a higher DR, but this is followed by a higher FPR and OAPR, whatever screening programme is used (Wald et al., 1999a). Comparing all screening methods for women older than 35 years, the probability of a positive result (giving the indication for amniocentesis) is specifically lower when using the integrated test (Wald et al., 1999a). Indeed, in the FASTER Trial, the screening performance of the integrated test for women aged 35 years or more was a DR of 91% for FPR of 2%. This should be compared to a DR of 92% and FPR of 13% for the quadruple test and a DR of 95% and FPR of 22% for the combined test (Malone et al., 2005a). In that view, the evidence suggests that women with advanced maternal age benefit most by integrated test screening.

**Serum marker levels**

Initial studies reported a significant difference in the level of second trimester biochemical markers between IVF patients and spontaneous pregnancies (Wald et al., 1999b). Evaluating 151 IVF pregnancies in which AFP, uE3, free hCG and total hCG were measured, median uE3 levels were 6% lower, median free hCG 9% higher and median total hCG 14% higher (all statistically significant) in IVF pregnancies compared with controls. These results might explain the higher FPR in IVF pregnancies, which is about twice as high as that in controls (Barkai et al., 1996). High hCG levels may be explained partly by a greater number of corpora lutea (Frishman et al., 1997), by multiple implantation sites or by progesterone supplementation, which increases placental hCG production. Alternatively, it may represent placentation failure that could result in changes in the trophoblast function and thus hCG production (Raty et al., 2003). The low uE3 levels remain unexplained. It was suggested that in DS screening in IVF pregnancies, hCG and uE3 values should be adjusted to avoid the high screen positive rate. A recent multi-centre study analysing 1515 singleton pregnancies by assisted reproduction techniques (ART) found a similar trend of 12% lower uE3, 7% higher hCG and 6% lower AFP, but these differences were not significant. In this study, the FPR did not differ from age-matched controls (Muller et al., 2003). This might suggest that the significance found by Wald et al. (Wald et al., 1999b) was a type I error. Other comparative studies show that serum levels of all three second trimester markers in ART pregnancies do not differ from controls (Rice et al., 2005).

The effect of ART on first trimester screening is also controversial. Some studies show an increase in hCG levels (Niemimaa et al., 2001) and a decrease in PAPP-A levels, which increase the FPR (in ART pregnancies) by an additional 1.2% (Liao et al., 2001). However, other study found no significant difference in serum levels of hCG, PAPP-A or FPR between ART and spontaneous pregnancies (Wojdemann et al., 2001). Interestingly, some recent studies found increased NT in ART pregnancies (MoM, multiples of median) (Maymon and Shulman, 2004; Hui et al., 2005).

The effect of ART on integrated (first and second trimester) screening was assessed in a group that underwent a serial disclosure DS screening programme (Maymon and Shulman, 2004). The rate of first trimester screening FPR was comparable to controls. The rate of second trimester FPR was two-fold higher in the ART group. It was concluded that ART singleton patients should be screened either by the combined or by the integrated tests (Maymon and Shulman, 2004). To conclude, it is postulated that DS screening in
Antenatal screening for Down's syndrome

ART pregnancies is associated with a higher FPR. To decrease the magnitude of such uncertain experience, it is advised to use better screening modalities, such as the combined or integrated tests.

Multiple pregnancies

The current concept is that first trimester NT measurement is superior to second trimester serum screening for multiple pregnancies (Maymon et al., 2005). All monochorionic pregnancies are monozygotic. On the other hand, most but not all dichorionic pregnancies are dizygotic. For a dizygous twin pregnancy, the risk of DS for each fetus is independent of the risk for the other. This implies that for dichorionic twin pregnancies, the pregnancy-specific risk is calculated by summing the individual risk estimates for each fetus (Meyers et al., 1997). On the other hand, in monochorionic pregnancies, the risk is based on an average of likelihood ratios derived from NT measurements of both twins. Therefore, diagnosis of chorionicity should be the first step in ultrasound evaluation of twins during the first trimester, and it has major implications on the noninvasive screening for aneuploidy in twins. First trimester scanning and measurement of NT enable the identification of a fetus-specific risk for Down’s syndrome in dichorionic pregnancies. In twin pregnancies, the sensitivity of fetal nuchal translucency thickness in screening for trisomy 21 is similar to that in singleton pregnancies, but the specificity in monochorionic pregnancies is lower because translucency is also increased in chromosomally normal monochorionic twin pregnancies (Sebire et al., 1996). In another small study, a DR of 100% was achieved with screen positive results of 4.3% (Maymon et al., 2001). The use of the first trimester combined test (Spencer and Nicolaides, 2003) can lower the FPR of NT measurement. It is estimated that by using ‘pseudo-risk’ and not specific risk figures, the integrated test can reach a DR of 78 and 93% for dichorionic and monochorionic twin pregnancies, respectively (Wald et al., 2003a).

Conclusion

Three major national studies (SURUSS, BUN and FASTER Trial) as well as many local studies (evaluated by Nicolaides, 2004) demonstrate that the first trimester combined test is equivalent or superior to second trimester serum screening (Simpson, 2005). The best test (with highest DR and lowest FPR) is the integrated test, which is also the safest (leading to lower rate of invasive procedures) and most cost-effective (Wald et al., 2004). We believe that current practice of screening should employ either the integrated test or the combined test for those women who want an early result and who are prepared to accept the higher risk of having an unnecessary invasive procedure. Stepwise sequential or contingent screening might be beneficial since not all women need to return for second trimester markers’ evaluation (Malone et al., 2005b), but these tests require implementation and cost-effectiveness studies before they can be recommended. The quadruple test should be reserved for women who only reach prenatal care in the second trimester. There is still limited data on the performance of these tests in ART pregnancies. Screening in twin pregnancies can be done only when chorionicity is certain. Twin pregnancies should be screened by NT measurement, combined or integrated tests, and it is not clear which test is superior.

References

B. Weisz and C.H. Rodeck


Submitted on January 27, 2006; resubmitted on March 21, 2006; accepted on March 21, 2006.