Minireview

Prenatal screening for trisomy 21: recent advances and guidelines

Jacob Canick*
Division of Medical Screening and Special Testing, Department of Pathology and Laboratory Medicine, Women and Infants Hospital, Alpert Medical School of Brown University, Providence, RI, USA

Abstract
The performance of prenatal screening tests for the identification of trisomy 21 (Down syndrome) has markedly improved since the 1970s and early 1980s when maternal age was the sole mode of screening the general pregnant population. With the discovery of second trimester serum markers in the 1980s and 1990s and implementation of double, triple, and quad marker testing; the discovery of first trimester serum and ultrasound markers in the 1990s and implementation of the combined test; and the development of the integrated test and sequential screening strategies over the past decade, the performance of screening has improved to a detection rate of 90%–95% at a false positive rate of 2%–5%. In this review, I will describe the advances in prenatal screening for trisomy 21, present current screening strategies, and discuss guidelines published by professional societies and regulatory bodies, with a focus on current prenatal screening practice in the USA.

Keywords: Down syndrome; multiple markers; prenatal screening; trisomy 21.

Introduction
Prenatal screening for Down syndrome, the most common cause of congenital developmental delay, has improved steadily during the last three decades. From the 1970s to 1980s, when maternal age alone was the major determinant of Down syndrome risk, the introduction of multiple markers measured in the late first and early second trimesters has provided screening performance that now reaches and exceeds 90% sensitivity and 95% specificity (90% detection rate and 5% false positive rate) (1, 2). From a historical perspective, prenatal screening as a population-based program began in the mid to late 1970s in the UK after the discovery that maternal serum α-fetoprotein (AFP) was, on average, elevated in pregnancies in which the fetus had an open neural tube defect (anencephaly or spina bifida) (3). The birth prevalence of the open neural tube defects in the UK was almost four per thousand (data from England and Wales) and the introduction of a rational, evidence-based screening program based on the measurement of maternal serum AFP was welcomed and well utilized. With the discovery in the 1980s that taking folic acid, beginning at, or before conception will prevent almost 80% of fetal open neural tube defects from forming, the combination of secondary prevention with AFP screening and primary prevention with folic acid has resulted in a reduction in the birth prevalence of these disorders to <0.3 per thousand (data from England and Wales).

Screening using AFP provided a new and elegant model for expressing and simplifying the quantitative aspects of risk assessment and identification of at-risk pregnancies. For the first time, the distribution shapes of a screening marker were determined for both affected and unaffected pregnancies, and gestational age-associated unaffected medians were established as the reference values, with the calculation of a patient-specific multiple of the median (MoM) as the normalized marker unit (3). Thus, the value 1 MoM became the center of the unaffected distribution for singleton pregnancies and, in the case of the open neural tube defects; an elevated MoM value became the center of each affected distribution. A screened woman was provided with her own MoM value which was then compared to the 1 MoM-centered, unaffected distribution and the higher MoM-centered, affected distribution. For open spina bifida, the median AFP value was found to be about 4 MoM, and for anencephaly, the median AFP value was found to be about 7 MoM. The degree of overlap between the affected and unaffected distributions determined the performance of that particular analyte as a screening marker. As the calculation of patient-specific risk became central to screening, the marker MoM value could be easily converted to a patient-specific likelihood ratio of having an affected pregnancy.

History of Down syndrome screening
It has been almost 30 years since Irwin Merkatz, speaking at a New York State Birth Defects Symposium, reported that maternal serum AFP levels were, on average, low in Down syndrome and Edwards syndrome pregnancies in the early second trimester (4). Cuckle and Wald quickly verified Merkatz’ finding and described how overlapping distributions of
Canick: Prenatal screening for trisomy 21

Article in press - uncorrected proof

Figure 1
MoM values, with medians, of the first and second trimester prenatal screening markers in 75 cases of Down syndrome. (A) The first trimester combined test markers, nuchal translucency (NT), free β-hCG, and PAPP-A. (B) The second trimester quad test markers, AFP, uE3, hCG, and inhA. Partial data from the FASTER Trial. MoM, multiple of the median.

Table 1
Screening performance (detection rate at a fixed 1% or 5% false positive rate) for the most common tests.

<table>
<thead>
<tr>
<th></th>
<th>SURUSS</th>
<th></th>
<th>FASTER</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detection rate, %</td>
<td>At 1% FPR</td>
<td>Detection rate, %</td>
<td>At 1% FPR</td>
</tr>
<tr>
<td>First trimester combined test</td>
<td>72</td>
<td>86</td>
<td>72</td>
<td>85</td>
</tr>
<tr>
<td>Second trimester triple test</td>
<td>56</td>
<td>77</td>
<td>45</td>
<td>69</td>
</tr>
<tr>
<td>Second trimester quad test</td>
<td>64</td>
<td>83</td>
<td>60</td>
<td>81</td>
</tr>
<tr>
<td>Serum integrated test</td>
<td>73</td>
<td>87</td>
<td>70</td>
<td>86</td>
</tr>
<tr>
<td>Full integrated test</td>
<td>86</td>
<td>94</td>
<td>87</td>
<td>95</td>
</tr>
</tbody>
</table>

Data from SURUSS (1) and FASTER (2), two large studies that were designed to compare screening markers across the first and second trimester in the same women. FPR, false positive rate.

Advances in screening: first trimester, integrated, or something in between

Screening performance estimates found in SURUSS and FASTER for the first trimester combined test (85% detection
rate for a 5% false positive rate) are somewhat lower than have been reported in some other studies. Whether or not first trimester screening reaches 85%, 90% or 95% detection rates for a fixed 5% false positive rate, it is important to note that the comparable integrated test screening performance will be higher if the cause of the improved performance (e.g., NT performs better in some hands) is shared by both tests. In Table 2, three different performance estimates for the first trimester combined test are shown in comparison to the expected changes that would follow in integrated test performance. It is unreasonable for a study to suggest that first trimester performance is the same as that of the integrated test if that study cannot provide the comparable integrated test performance.

During the past decade, as an outgrowth of the integrated test, studies were published indicating that two step screening when correctly applied, could provide both early and high performance screening (8–10). In its simplest form, a woman would expect to have the complete integrated test with her risk reported when all markers have been measured. If, however, after her first trimester markers had been measured, her risk of Down syndrome pregnancy based on those markers was found to be higher than an especially high risk cut-off, it would be reasonable to “triage” that woman into a special high risk category and counsel on the risks and benefits of early diagnostic testing. It would be expected that no more than about 1% of all screened women would fall into this very high risk group, and that the 99% of screened women below the high risk threshold would continue on to the second stage of the integrated test. At least 60% of all the Down syndrome cases would be identified in the 1% of women triaged early, and an additional 25%–30% of the Down syndrome cases would be identified in the 1%–2% of women identified as high risk by the complete integrated test. In this way, a sequential strategy would provide almost as good screening performance, in terms of low false positive rate, as the integrated test, with the added benefit of identifying the majority of cases early. A comparison of expected screening performance of the first trimester combined test, the integrated test, and a sequential strategy, published by Palomaki et al. in 2006 (9) is shown in Table 3. When applying a sequential screening strategy, the benefits of improved screening performance are optimized by keeping the risk cut-off at the first and second stages of screening high, resulting in a smaller overall false positive rate. When the cut-off chosen for the first trimester stage is not set high enough, the stage one false positive rate will be too great, and the test will provide almost no improvement in performance over what can be achieved with the first trimester combined test alone.

In addition to sequential screening protocols, contingent protocols have been proposed, in which three categories of risk are reported after the first trimester markers are measured (11). The three risk categories are 1) a very high risk group which is offered early prenatal diagnosis, 2) a low risk group which is considered screen negative with no further action recommended, 3) and a group with risks intermediate between the high and low cut-offs which is recommended to continue to the second stage of the integrated test with risk reporting (screen positive and screen negative) in the second trimester. A benefit of the contingent protocol is the early identification of a new low risk group requiring no further testing, that could comprise as many as 75% of all women screened. A drawback of the contingent protocol is its increased complexity, with a total of five different risk group identified, and the identification of a large intermediate group (comprising approx. 25% of all women screened) who are effectively alerted that they have an interim positive result and would benefit from further testing. Clinical studies to test the effectiveness of and patient satisfaction with contingent screening are needed.

The laboratory’s role in nuchal translucency measurement

Arguably, the most informative of the current prenatal screening markers is fetal NT, a marker that is measured in the ultrasound unit, not the clinical laboratory. Yet, in many screening programs the laboratory must incorporate the quantitative data that come from the sonographer into a multiple marker-based risk calculation. Therefore, the laboratory must assume some level of responsibility for the suitability of the NT measurement, its conversion to a normalized unit, and its quality assurance.

The sonographer’s task of measuring NT correctly is not the responsibility of the laboratory. Professional groups, like

<table>
<thead>
<tr>
<th>Test</th>
<th>False positive rate, %</th>
<th>Detection rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>At 84.3% DR</strong></td>
<td><strong>At 2.0% FPR</strong></td>
</tr>
<tr>
<td>First trimester combined</td>
<td>6.0</td>
<td>74.5</td>
</tr>
<tr>
<td>Sequential</td>
<td>2.0</td>
<td>84.3</td>
</tr>
<tr>
<td>Full integrated</td>
<td>1.2</td>
<td>87.4</td>
</tr>
</tbody>
</table>

DR, detection rate; FPR, false positive rate. Data from Palomaki et al. (9).
Figure 2 Measured values of (A) second trimester maternal serum AFP and (B) first trimester nuchal translucency (NT) from unaffected pregnancies as a function of increasing gestational age. Both AFP and NT increase in a log-linear fashion with increasing gestational age, AFP at approximately 15% per week and NT at approximately 20% per week. Note that on a log scale of measured values, both AFP and NT conform closely to Gaussian distributions. MoM, multiple of the median. AFP measured in international units per milliliter (IU/mL) and NT measured in millimeters (mm).

Figure 3 Log-Gaussian distributions of MoM values for (A) first trimester nuchal translucency (NT) and (B) second trimester hCG in unaffected and Down syndrome pregnancies. The unaffected NT MoM distribution, with a standard deviation (SD) of 0.11, is much tighter than the unaffected hCG MoM distribution, which has a SD of 0.24 (SD’s calculated from the log MoM values of each marker). The Down syndrome distributions of NT and hCG are similar, with both centered at 2.0 MoM and having almost the same SD (approx. 0.25). By showing the screening performance of each marker at a ≥2 MoM cut-off, this Figure demonstrates that the 50% detection rate provided by each marker is associated with a much lower false positive rate for NT (1%) than for hCG (8%).
established for each laboratory for the assay method being used. For NT, medians ideally should be established for each operator, i.e., each sonographer’s NT measurements with matching crown rump length measurements (to calibrate the gestational age at the time of the ultrasound) are used to calculate that sonographer’s own medians. When many sonographers attempt to conform to a single, mandated set of NT medians, it puts a burden on the operators that is unnecessary and, in documented studies, not possible. Studies from the Netherlands, Scotland, and the USA underscore the difficulty that sonographers can have in attaining a standard set of medians in their practice (13–15). In one of these studies, data are presented that clearly support the implementation of operator-specific medians to yield more consistent screening performance for NT.

**Prenatal screening policies and guidelines**

Given the variety of tests currently available, both governmental and professional groups in a number of countries have published guidelines on prenatal screening for Down syndrome. The consensus appears to be to provide the best screening performance, with early screening stressed by some.

In Europe, based on data from the EUROCAT participants (the European Surveillance of Congenital Anomalies consortium), early screening is preferred, with first trimester combined screening recommended in eight of the nine EUROCAT countries that have a national screening policy (16). In six of the nine, triple or quad marker testing is recommended when a woman presents for prenatal care after 13 weeks. In 2007, the UK, a EUROCAT country, published performance guidelines that called for a 90% detection rate at <2% screen positive rate by April 2010, performance achievable using the integrated test or sequential variants (17). However, the UK National Screening Committee continues to recommend first trimester combined screening as the national standard. Because the 2010 performance target is acknowledged to be unattainable using the first trimester combined test, current UK performance guidelines continue to be the 2007 standard, a detection rate >75% at a screen positive rate of <3%.

In Canada and the US, the major obstetrics and gynecology societies, the Society of Obstetricians and Gynaecologists of Canada (SOGC) and the American College of Obstetricians and Gynecologists (ACOG) separately published screening guidelines early in 2007 (18, 19). The SOGC guidelines state that “in 2007, as a minimum standard, any prenatal screen should have a 75% detection rate with no more than a 5% false positive rate for Down syndrome,” with the following tests meeting that standard: first trimester screen, second trimester quad screen, and two step screens (contingent, integrated, serum integrated, sequential). The ACOG guidelines do not specify a minimum level of screening performance. Rather, they directly state that:

- “patients seen early in pregnancy should be offered … screening that combines first- and second-trimester testing (integrated or sequential).”
- “the screening strategy chosen will depend on the availability of CVS and of personnel trained in NT measurement.”
- “when CVS is not available, … offer integrated screening to patients who present in the first trimester.”
- “if NT measurement is not available or cannot be obtained … offer serum integrated screening to patients who present early and second-trimester screening to those who present later.”

Both SOGC and ACOG include strongly worded statements on revising the role of maternal age in prenatal screening for Down syndrome. The Canadian guidelines state that, “maternal age screening is a poor minimum standard for prenatal screening for aneuploidy and should be removed as an indication of invasive testing.” The US guidelines state “all women should be offered aneuploidy screening … regardless of maternal age.” It is now agreed that while maternal age should be used in multiple marker risk estimation, it is no longer considered the primary determinant of risk.

**Conclusions**

It is now possible, using multiple marker methods, to detect more than 90% of cases of fetal Down syndrome in the 5% or less pregnant women assigned as screen positive. This level of performance can be viewed both in positive and negative terms. On the positive side, it is quite an achievement to have a screening test with such high sensitivity (≥90%) and specificity (≥95%). On the negative side, most of the 2%–5% of pregnant women who are found to be screen positive will be false positive, but will still be offered an invasive and risky diagnostic procedure for chromosome analysis.

To put this in a quantitative context, if we consider the background risk of Down syndrome pregnancy to be 1 in 500, applying a screening test with a 90% detection rate and a 5% false positive rate to this population means that 1 in every 28 women called screen positive will actually have an affected pregnancy, while 27 of the 28 will have been falsely called positive. If the screening test has a 90% detection rate and only a 2% false positive rate, one in every 11 women called screen positive will actually have an affected pregnancy, while 10 of the 11 will have been falsely called positive.

To achieve the level of performance possible today, serum markers must be combined with the fetal ultrasound marker, nuchal translucency, requiring close communication between the clinical laboratory and the obstetrical ultrasound unit. In practical terms and in terms of equity in screening, there must also be widespread availability of fetal sonography to pregnant women. In countries in which most of the population is in close proximity to appropriately trained ultrasound units, equity can come close to being achieved. In countries with dispersed populations, however, a large proportion of pregnant women may not have the opportunity to receive the best screening tests. This is a dilemma when screening
includes specialized patient testing such as sonography, but will be improved if and when new screening methods that exclusively use laboratory tests are developed. Efforts to find new markers, including the possibility of circulating free DNA in the maternal circulation, are currently underway.

Conflict of interest statement

Authors’ conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

References