INTRODUCTION

Medical screening programs in general are not intended to provide diagnostic testing results, but rather to detect populations at increased risk for a particular disorder or condition. Unfortunately, in most cases, the screening test values obtained for affected individuals significantly overlap those from the unaffected population. In general, screening parameters should be set so that a majority of the affected cases are detected, while the number of false positives is kept as low as possible. For example, as discussed in the FP-C Participant Summary for 2004, screening for open neural tube defects using maternal serum Alpha-fetoprotein (AFP) has generally been set to detect essentially all fetuses with anencephaly and approximately 75 percent of those with open spina bifida, with a false positive rate of less than 5 percent (recently circa 2 percent).^1

It was recognized in the late 1980s that second-trimester maternal serum levels of AFP were lower in women carrying infants with Down syndrome (DS), and that combining the risks associated with age and with different AFP levels could greatly increase the prenatal detection of affected fetuses, while reducing the need for amniocentesis in pregnant women over the age of 35.\textsuperscript{2} However, the proposed risk cutoffs were expected to detect only 40 percent of affected pregnancies, while 7 percent of screened women would need amniocentesis, versus 11 percent using maternal age alone. Many programs began screening with risks calculated from AFP levels plus maternal age, while others preferred to continue recommending amniocentesis in all pregnant women over age 35 because of the lower detection rate in this population using the recommended risk cutoffs with the addition of AFP.

The subsequent addition of other markers has offered the possibility of much higher detection rates (DR) for DS, lower false positive rates (FPR), or both, depending upon the markers and cutoffs chosen. Several other factors are important as well, as discussed below.

MULTIPLE MARKER SCREENING FOR DS

Over the last 15 years, several additional markers for second trimester DS screening have been proposed, including, among others, unconjugated estriol (uE\textsubscript{3}); free β, total, or α human chorionic gonatrophin (hCG); and dimeric inhibin-A (DIA), which were proposed more-or-less in that order.\textsuperscript{3} Each addition increased calculated detection rates, decreased the number of patients requiring further testing, or both. However, because of differences in cutoffs, laboratory analytical performance, and distribution of maternal ages of screened populations, accurate comparison of performance of various combinations of markers has been impossible until recently.

In 2003, Wald and others reported the results of a prospective study of 47,053 singleton pregnancies (101 with fetal DS) from 25 maternity units, the so-called Serum Urine and Ultrasound Screening Study (SURUSS).\textsuperscript{4} The study was designed to evaluate testing parameters for both first and second trimester screening, as well as various combinations of the two. The current CAP survey focuses only on second trimester screening (14-20 completed weeks), so only the results for this will be summarized here.
As shown in the table, the use of quadruple screening (AFP, uE3, dimeric inhibin –A (DIA), and either total or free β hCG, along with maternal age) effectively doubles the DR for DS with a constant FPR of 5 percent; alternatively, for a DR of 85 percent, the FPR is reduced by > 80 percent (40 percent for AFP alone vs. ~7 percent for quadruple).

Table. AFP vs. Multiple Marker Screening for DS at 14-20 completed weeks.*

<table>
<thead>
<tr>
<th>Mat. Age +</th>
<th>DR in percent for FPR of 1 percent</th>
<th>DR in percent for FPR of 3 percent</th>
<th>DR in percent for FPR of 5 percent</th>
<th>FPR in percent for DR of 70 percent</th>
<th>FPR in percent for DR of 75 percent</th>
<th>FPR in percent for DR of 80 percent</th>
<th>FPR in percent for DR of 85 percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>19</td>
<td>33</td>
<td>42</td>
<td>21</td>
<td>26</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>AFP + t-hCG</td>
<td>40</td>
<td>57</td>
<td>66</td>
<td>6.4</td>
<td>8.6</td>
<td>11.7</td>
<td>16</td>
</tr>
<tr>
<td>AFP, t-hCG, uE3</td>
<td>51</td>
<td>67</td>
<td>74</td>
<td>3.7</td>
<td>5.2</td>
<td>7.4</td>
<td>10.9</td>
</tr>
<tr>
<td>AFP, t-hCG, uE3, DIA</td>
<td>62</td>
<td>75</td>
<td>83</td>
<td>2.0</td>
<td>2.9</td>
<td>4.5</td>
<td>7.1</td>
</tr>
</tbody>
</table>

*Modified from reference 4.

Abbreviations: DR, detection rate; FPR, false positive rate; AFP, α-fetoprotein; t-hCG, total human chorionic gonadotrophin; uE3, unconjugated estriol; DIA, dimeric inhibin-A

Some aspects of the study have been criticized; as the authors themselves acknowledge, the outcome data may be slightly biased, for example. Regardless of this, it is obvious that multiple marker screening has significant advantages regarding DR and FPR. The major negative factor in quadruple screening is the cost of testing, although the benefit to the total health budget of performing quadruple testing is high.

OTHER FACTORS

In the SURUSS study, adjustment for maternal weight had only a very small beneficial effect on the overall detection rates (~0.2 percent). However, adjustment definitely tends to “even out” the DRs and FPRs over the range of maternal weights, particularly at the extremes. The use of ultrasound dating reduced the FP rates by about 20 percent, as compared to last menstrual period (LMP) dating alone. Adjustments for maternal weight, ultrasound dating, and racial/ethnic group have been discussed in depth in recent CAP surveys.

Increasing assay precision can affect both DR and FPR by effectively separating the distributions of affected and unaffected populations. Assay precision is taken into account in most if not all risk calculation programs, although many use fixed precision levels rather than those for the assays being used in a given laboratory. The Wald 2000 parameter set, for example, uses the assay precisions obtained with the Perkin Elmer AutoDelphia® (Perkin Elmer Life Sciences, Boston, Mass), which has shown the best precision values (i.e., the lowest imprecision, or CVs) for AFP, uE3, and hCG in recent CAP surveys.
The maternal serum concentrations of the screening analytes are also affected by many other things, including fetal gender. For example, female gender is associated with higher maternal serum hCG and lower AFP levels than those seen with male fetuses. However, in a recent study, no gender-related differences were seen in DS-affected fetuses, and the false negative rates were not affected by correcting for gender.\(^7\) Also, women with a history of fetal aneuploidy—specifically trisomy 13, 18, or 21—tend to have significantly higher hCG and pregnancy-associated plasma protein-A (PAPP-A) levels in subsequent pregnancies.\(^8\) Some programs correct for history of a DS pregnancy now, but more extensive correction may be indicated in the future.

**MONITORING PERFORMANCE**

It should be obvious that an essential component of a maternal serum screening program is that each laboratory must assess its assay performance continually. In addition to running calibrators and controls, monitoring the median MoM (Multiples of Median) values for each analyte on a daily or weekly basis gives a very good estimate of assay drift—i.e., gradual increases or decreases in assay values obtained. Minor fluctuations are to be expected, since the patient sample mix varies (the mixture of gestational ages, maternal ages, racial/ethnic groups, and so forth). However, a continued trend requires investigation.\(^9\) External proficiency programs such as this CAP survey also evaluate laboratory assay performance, but on an occasional “spot” basis.

At the same time, the ultimate concern of a screening program is the detection of disorders that are clinically important and that lend themselves to some type of effective intervention. Thus assessment of a program involves determining how well the test(s) discriminate between affected and unaffected pregnancies and what the odds of having an affected fetus are for those with a positive screening result.\(^10\)

The easiest method of epidemiological evaluation is by checking the initial positive rate (IPR) and the median MoMs for each assay on a regular basis.\(^5\) In the case of DS screening, the maternal age distribution should be considered in evaluating IPR, since women over 35 years would be expected to have more positive screens.\(^11\) However, it is more important to determine the detection rate and odds of having an affected fetus; this requires obtaining outcome results on screened pregnancies. The DR and IPR together can be used to calculate false negative and positive rates.

The odds of having a positive result (OAPR), in this case an affected fetus, given a positive screening result is the ratio of true positives to false positives. For example, if 100 women have positive screens but only 20 have an affected fetus, the OAPR would be 20:80, or 1:4.

Estimates of all of these measures for a given combination of screening tests are available in the literature, but good practice demands that they be confirmed, insofar as possible, in individual programs.
Summary

The effective use of multiple markers for Down syndrome screening — for example, α-fetoprotein, unconjugated estriol, β-hCG, and dimeric inhibin A — can both increase the detection rate and decrease the false positive rate. Other factors, including the level of assay imprecision and adjustment for maternal weight and ethnic group, also can influence screening efficiency.

Monitoring of a screening program should include:

a. continual assessment of assay performance by following median MoM values,
b. participation in an external proficiency testing program,
c. frequent assessment of initial positive rates (IPR), and
d. periodic determination of actual detection rates, as well as false positive and false negative rates.

For the last of these – assessment of epidemiological performance – it is necessary to monitor outcomes of the screened pregnancies.
References


