First-Trimester Risk Calculation for Trisomy 13, 18, and 21: Comparison of the Screening Efficiency between 2 Self-Developed Programs and Commercial Software

Steen Sørensen, Günther Momsen, Karin Sundberg, Lennart Friis-Hansen, and Finn Stener Jørgensen

BACKGROUND: Reliable individual risk calculation for trisomy (T) 13, 18, and 21 in first-trimester screening depends on good estimates of the medians for fetal nuchal translucency thickness (NT), free β-subunit of human chorionic gonadotropin (hCGβ), and pregnancy-associated plasma protein-A (PAPP-A) in maternal plasma from unaffected pregnancies. Means and SDs of these parameters in unaffected and affected pregnancies are used in the risk calculation program. Unfortunately, our commercial program for risk calculation (Astraia) did not allow use of local medians. We developed 2 alternative risk calculation programs to assess whether the screening efficacies for T13, T18, and T21 could be improved by using our locally estimated medians.

METHODS: We established these estimates from 19,594 women with singleton pregnancies and from 100 pregnant women carrying a fetus affected with trisomy (11 with T13, 23 with T18, and 66 with T21). All measured values were recalculated to a multiple of the median (MoM) and log10 transformed; the mean and SD were calculated for each group.

RESULTS: At a given risk cutoff value, we observed a slight improvement in detection rate (DR) for T13, T18, and T21 for a slightly higher false-positive rate (FPR) compared with the commercial program. The lower FPR in the commercial program was mainly caused by an inaccuracy in the PAPP-A median.

CONCLUSIONS: Center-specific medians for NT, hCGβ, and PAPP-A should be used in risk calculation programs to ensure high DRs and low FPRs for all 3 trisomies at a given risk cutoff.

© 2011 American Association for Clinical Chemistry
Materials and Methods

STUDY PARTICIPANTS
We collected data from women with singleton pregnancies who attended first-trimester screening for Down syndrome at 2 centers (9775 at Hvidovre Hospital and 9819 at Rigshospitalet). Approval was given by the Danish National Board of Health. The blood samples were collected between July 1, 2005, and June 30, 2007. We extracted nonidentifying maternal demographic information and ultrasound and biochemical determinations from a clinical database and software (Astraia v 1.17.68, Siemens). Only pregnant women with a single living fetus, no trisomy, and a complete data set of ultrasound and biochemical measurements were included. The mean maternal age was 30.4 years (range 16–45 years); 16.5% of the women were ≥35 years old. We performed no correction for weight or smoking, as we did not record these parameters at that time for all pregnant women.

The study also included 100 pregnant women carrying a fetus with trisomy, with a complete data set identified in the 2 centers from a 4-year period, 11 with T13 [median age 32 years (range 27–41)], 23 with T18 [32 (25–43) years], and 66 with T21 [34 (23–44) years]. All true positives and false negatives in the screening during this period were included.

ULTRASOUND MEASUREMENTS
Sonographers certified to measure NT thickness by the Fetal Medicine Foundation, UK, performed the ultrasound examinations. The gestational age was obtained by measuring the crown-rump length (CRL); 16–82 mm corresponding to a gestational age from 8 weeks and zero days (8 + 0) to 13 weeks and 6 days (13 + 6) (5), as calculated by Astraia. We measured NT thickness only when the CRL was at least 45 mm. If it was below this value, the woman was given a new appointment for an ultrasound examination at a more appropriate gestational age.

BLOOD SAMPLING
The blood samples were collected at outpatient clinics in serum separator tubes containing a polymer gel and clot activator (Becton Dickinson), very often 1–2 weeks preceding the ultrasound investigation. After collection, the samples, which had clotted at room temperature, were centrifuged at 1850g for 10 min, and the serum was stored at −20 °C until assayed. After testing, the samples were stored at −80 °C in case of retesting.

BIOCHEMICAL MEASUREMENTS
We measured the concentrations of PAPP-A and hCGβ by use of a homogeneous random-access semi-automated assay, employing time-resolved amplified cryptate emission (TRACE) technology using the Kryptor instrument (Brahms AG). We ensured analytical accuracy throughout the study by having 3 specimens analyzed monthly by the UK National External Quality Assessment Service. When the final report returned, we related the hCGβ and PAPP-A values to the mean values for all Kryptor users. The bias was slightly positive for hCGβ and slightly negative for PAPP-A, overall less than 5% during the 2-year study period (see Supplemental Table S1, which accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue7).

STATISTICAL ANALYSIS
We performed statistical investigations and curve fitting using Prism version 4 (GraphPad). To account for multiple comparisons ranging from 56 to 97 gestational days, we used an adjusted statistical significance of P < 0.001, corresponding to a nonadjusted statistical level of significance of P < 0.05.

RISK CALCULATION
We used the commercial software program, Astraia, for the routine risk calculation of T13/18 and T21 during the first trimester. If the risk reached or exceeded a risk of 1 in 300 for T21 or 1 in 150 for T13/T18, the pregnant women were routinely offered a definitive diagnosis by chorionic villus sampling or amniocentesis for fetal karyotyping.

We developed 2 risk-calculation programs using Microsoft Office Excel 2003, the first 1-dimensional and the second 3-dimensional. Acquired and measured values from each pregnant woman were entered into the appropriate columns in an Excel worksheet as follows: maternal age, CRL, NT, gestational age in weeks + days (from the CRL), hCGβ, PAPP-A, and calculated values as follows: \( \log_{10} NT \) (MoM), \( \log_{10} hCG\beta \) (MoM), and \( \log_{10} PAPP-A \) (MoM).

We calculated the likelihood ratio, LR, as the ratio of probability densities from 2 comparable 1-dimensional frequency density distributions (gauss-
ian) of log$_{10}$ values (MoM) in affected, $f_1(x)$, and unaffected, $f_2(x)$, pregnancies:

$$LR = \frac{f_1(x)}{f_2(x)}.$$  

The age-related (a priori) maternal age risk, $P_{pre}$, for T13, T18, and T21, which rises with increasing maternal age, was calculated by means of 3 exponential equations using data on the prevalence of T13, T18 (6), and T21 (7) at a fixed gestational age of 12 weeks (see online Supplemental Table S2).

The posttest risk for T13, T18, or T21 could now be determined at the time of the test by:

$$P_{post} = LR(NT) \times LR(PAPP-A) \times LR(hCG) \times P_{pre}$$

assuming independence between the variants, that is, the absence of correlation between them (naive Bayes).

For multivariate normal distribution, we organized the data into an $n \times p$ matrix, $X$, where the $p$ columns were measurements and calculated values, and the $n$ row vectors were the 3 attributes: log$_{10}$ NT (MoM), log$_{10}$ hCGβ (MoM), and log$_{10}$ PAPP-A (MoM). Covariance matrices, $S$, were calculated for all sets of data. We used the Jacobi transformation (8) to calculate the inverses, $S^{-1}$, and the determinants, $|S|$, of the covariance matrices. The probability densities (likelihoods, $f$) were calculated by the 3-dimensional multivariate normal density distribution. In the case of a trivariate distribution, the corresponding formula is:

$$P_{post} = LR(NT, PAPP-A, hCG) \times P_{pre}.$$  

See online Supplemental Notes for a more detailed explanation of the multivariate normal distribution.

### Results

**NUCHAL TRANSLUCENCY THICKNESS MEASUREMENTS**

We determined the median values of NT for CRL from 45–82 mm for nonaffected pregnancies. The differences between the median values of the 2 centers were, on average, zero, and the range was from −0.2 to 0.2. A second-order polynomial equation, found by the regression method, described the local median curve: NT (median) = $B_0 + B_1 \cdot x + B_2 \cdot x^2$, where $x$ was the

### Table 1. Log$_{10}$ NT (MoM), log$_{10}$ hCGβ (MoM), and log$_{10}$ PAPP-A (MoM) for unaffected pregnancies and pregnancies affected by T13, T18, or T21.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Unaffected</th>
<th>T13</th>
<th>T18</th>
<th>T21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Log$_{10}$ NT (MoM)$^a$</td>
<td>0.00</td>
<td>0.10</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.12</td>
<td>0.47</td>
<td>0.25</td>
</tr>
<tr>
<td>Log$_{10}$ hCGβ (MoM)</td>
<td>−0.03</td>
<td>0.26</td>
<td>−0.44</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.25</td>
<td>−0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>Log$_{10}$ PAPP-A (MoM)</td>
<td>−0.01</td>
<td>0.24</td>
<td>−0.36</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.23</td>
<td>−0.36</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.25</td>
<td>−0.37</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.24</td>
<td>−0.70</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.22</td>
<td>−0.52</td>
<td>0.25</td>
</tr>
</tbody>
</table>

T13/T18

<table>
<thead>
<tr>
<th>Marker</th>
<th>Unaffected</th>
<th>T13</th>
<th>T18</th>
<th>T21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Log$_{10}$ NT (MoM)</td>
<td>0.00</td>
<td>0.12</td>
<td>0.45</td>
<td>0.25</td>
</tr>
<tr>
<td>Log$_{10}$ hCGβ (MoM)</td>
<td>0.00</td>
<td>0.24</td>
<td>0.70</td>
<td>0.30</td>
</tr>
</tbody>
</table>

$^a$ NT (MoM) is not used in Astraia.

$^b$ Values from own program.

$^c$ Values from Astraia.
measured CRL (see online Supplemental Fig. S1). Estimates gave $B_0 = -1.73$, $B_1 = 0.08483$, $B_2 = -0.0004904$. We converted all NT values to $\log_{10} NT$ (MoM) by dividing the measured value by the expected NT median of unaffected pregnant women from the CRL vs the NT regression curve and then taking the logarithm of this value. The normal probability plot of $\log_{10} NT$ (MoM) values, from 1st to 99th percentiles, fitted a reasonably straight line from $-0.23$ to $0.23$ (see online Fig. S2A), indicating a gaussian distribution in this range. Values below the 1st percentile or above the 99th percentile could indicate an abnormality of a different kind. The plot did not pass a normality test (D’Agostino–Pearson omnibus normality test, $P < 0.001$). However, subsequent statistical analyses assumed log normality. The means and SDs of $\log_{10} NT$ (MoM) of the affected and nonaffected populations are given in Table 1. From these values, we generated 4 frequency distribution curves (gaussian), 1 for unaffected pregnancies and 1 each for T13, T18, and T21 pregnancies. For a given $\log_{10} NT$ (MoM) value, 3 quotients, LRs, could now be calculated from the height of the distribution curve for affected pregnancies divided by the height of the distribution curve for unaffected pregnancies. Very low and very high NT MoM values gave extremely high LR results, forming a nearly U-shaped curve (see online Supplemental Fig. S2B).

DOUBLE TEST

The measurements of hCG and PAPP-A were mainly between gestational day 63 and day 95, with at least 50 samples or more per day from each center, peaking on day 77 with nearly 600 samples (see online Supplemental Table S3). Medians for hCG and PAPP-A were estimated for all gestational days from day 56 to day 97 for each center and combined (see online Table S3). PAPP-A and hCG were at the limit to be significantly different on 3 of 35 and 1 of 35 gestational days, respectively ($0.001 < P < 0.01$). The difference in the means for each gestational day between Hvidovre Hospital and Rigshospitalet was, on average, 2% (range −19% to 31%) for hCG and −7% (range −27% to 17%) for PAPP-A (see online Supplemental Table S3). The largest differences were in the first 2 weeks, when the fewest samples were available. When pregnant women at Hvidovre Hospital were compared with those at Rigshospitalet of just 1 gestational day earlier, there were only 3 days with insignificant differences ($0.01 < P < 0.05$). The median values from each center were not randomly scattered around the default Astraia median curves. Both hospitals had hCG and PAPP-A median values above the Astraia median at very early gestational days and below it at later gestational days of the first trimester (Fig. 1). Thus, the slope of the default median curve was too low for hCG and too high for PAPP-A. Two fourth-order polynomial equations described the local smoothed median curves (see online Supplemental Table S4), with the shortest distances from the median values to the median curves (Fig. 2, A and C) also illustrated in the residual plots (Fig. 2, B and D); most of the values are within 5% from the curve.

All measured hCG and PAPP-A results were expressed in MoM values for the appropriate gestational day from the median curves described by the fourth-order polynomial equations (see online Supplemental Table S4), and transformed to $\log_{10}$ values. When comparing $\log_{10}$ hCG (MoM) values obtained from the Astraia program with those from our own program, the differences were small (Fig. 3, A and B). But for $\log_{10}$ PAPP-A (MoM) using the commercial program, we found high values between day 61 and day 76 (Fig. 3C), which were reduced to approximately zero by our own program (Fig. 3D).

We calculated the mean and SD of $\log_{10}$ hCG (MoM) and $\log_{10}$ PAPP-A (MoM) for unaffected and affected pregnancies (Table 1). We generated frequency distribution curves, 4 for $\log_{10}$ hCG (MoM)
(1 each for T13-, T18-, and T21-affected pregnancies and 1 for unaffected pregnancies) and 4 for log_{10} PAPP-A (MoM) (1 each for T13-, T18-, and T21-affected pregnancies and 1 for unaffected pregnancies).

The height of the affected curve divided by the height of the unaffected curve (LR) increased for T13 and T18 when hCG/H9252 (MoM) decreased and increased for T21 when hCG/H9252 (MoM) increased (see online Supplemental Fig. S3A). Curve lines occurred because the SDs were different for the unaffected and affected populations. When PAPP-A (MoM) decreased, the LR increased for all 3 trisomies (see online Supplemental Fig. S3B).

**RISK CALCULATION**

We compared DR and FPR at different risk cutoffs for T13, T18, and T21 using the Astraia software with those obtained with our own programs, assuming either univariate or trivariate normal distribution.

By our own calculation program, using univariate normal distribution and a risk of 1 in 150, we identified 8 of 11 (72.7%) T13 cases and 21 of 23 (91.3%) T18 cases for FPRs of 1.02% and 1.45%, respectively; at a risk of 1 in 300, we identified 87.9% of T21 cases for FPR of 4.7% (Table 2).

Univariate risk calculation assumes noncorrelation between the parameters. However, weak significant correlations exist between log_{10} NT (MoM) and log_{10} hCGβ (MoM) in unaffected pregnancies, and between log_{10} hCGβ (MoM) and log_{10} PAPP-A (MoM) in unaffected as well as T21-affected pregnancies (see online Supplemental Table S5). This correlation can be taken into consideration, if we use a trivariate normal distribution for these 3 parameters (4). If we set the risk at or above 1 in 100, DR for T13 was 54.5% for FPR of 0.55% (Table 2). However, 2 cases with a “T21 pattern” were detected and DR increased from 54.5% to 72.7%.

Univariate risk calculation assumes noncorrelation between the parameters. However, weak significant correlations exist between log_{10} NT (MoM) and log_{10} hCGβ (MoM) in unaffected pregnancies, and between log_{10} hCGβ (MoM) and log_{10} PAPP-A (MoM) in unaffected as well as T21-affected pregnancies (see online Supplemental Table S5). This correlation can be taken into consideration, if we use a trivariate normal distribution for these 3 parameters (4). If we set the risk at or above 1 in 100, DR for T13 was 54.5% for FPR of 0.55% (Table 2). However, 2 cases with a “T21 pattern” were detected and DR increased from 54.5% to 72.7%. DR for T18 was high, 91.3%, for FPR of 0.84% at a risk of 1 in 100. The Astraia program gave a combined DR for T13 and T18 of 76.5% for FPR of 0.53% at a risk of 1 in 150 (Table 2).
If we use the same risk for T21 as for T13 and T18 (1 in 100), DR will be 84.8% for FPR of 2.4%. DR can be increased to 89.4% by accepting a higher FPR of 4.55% at a risk of 1 in 250; this means that to get a 5% better DR, FPR must be doubled (Table 2). In Astraia, DR for T21 is lower at the same risk, 84.8%, for a lower FPR of 2.9%. The lower PAPP-A median in the Astraia program from day 56 to 76 reduced FPR but not DR, probably because of few data (Table 3).

Discussion

We compared the medians for NT, hCGβ, and PAPP-A for each gestational day from day 56 to 97 between 2 centers, which had identical clinical setup, sample handling, and analytical platform protocols. We observed no differences in the NT medians, but small differences were noted for hCGβ and PAPP-A, on average 2% for hCGβ and 7% for PAPP-A. The reason for these differences is unclear but might be due to a combination of analytical and CRL bias. Gestational age was calculated using the same version of a commercial program, Astraia, based on a given CRL formula (5). However, the mathematical function underestimates the gestational age by 1–2 days (9) and, recently, a slight modification of the formula was recommended to eliminate the underestimation (10).

Comparing our medians for hCGβ and PAPP-A with the default medians in the commercial software, we found our hCGβ and PAPP-A median curves to be steeper and flatter, respectively. This means that the MoM values from commercial software are increased in the early gestational period of the first trimester and decreased in the late gestational period of the same trimester. In this respect, the risks counteract each other. However, the risk is mainly driven by changes in PAPP-A and, to a minor extent, hCGβ. This gives a lower FPR for T21 (about 1.5%) for weeks 8–10 compared with weeks 11–13 using the Astraia program because of an inaccuracy of the hCGβ and PAPP-A medians, particularly at a lower gestational age. A similar trend was observed in another study but was found to be nonsignificant (2). However, we had no difference in DR in weeks 8–10 between our own programs based on univariate or trivariate normal distribution and the Astraia program, probably because of the rather small number, 25 of 66 T21 cases, in this gestational period. A simulation study found that 10% bias for a marker MoM value can change FPR 1% to 2% (11). To avoid incorrect medians, the software program should have
locally derived medians with a possibility of rolling median calculation (12), or at least medians that are appropriate for the screened population, as recommended in guidelines from the American College of Medical Genetics (13). In the new version of the Astraia software, laboratories can adjust the medians by 15% but cannot correct the slopes. This is insufficient and unsatisfactory.

The distribution parameters in this study are essentially in agreement with those published (Table 1), although some differences exist for mean of \( \log_{10} \text{hCG} \) (MoM) in T21 and mean and SD of \( \log_{10} \text{PAPP-A} \)
Data in the upper tail of the log10 NT distribution do not give an additional age when the NT is at least 0.8 mm, or if values can be avoided by measurement of NT at a later gestational age. To avoid an unreliable high-risk estimate, truncation limits have been used, especially for NT, to eliminate values outside set limits giving an improbable high LR. Suggested values have been 0.5 (MoM) and 2.5 (MoM) (14) and 0.8 (MoM) (15). Values outside the limits were set to the nearest limit for the purpose of risk calculation. In this study, only 22 results were < 0.45 MoM, giving a LR for T21 of 3 or higher. This can be avoided by measurement of NT at a later gestational age when the NT is at least 0.8 mm, or if values below 0.8 mm are set to 0.8 mm. Of the very high NT values, 40 results were > 2.5 MoM, giving a LR > 2000. Data in the upper tail of the log10 NT distribution do not fit a gaussian distribution beyond a value of 0.24, corresponding to above 99%, some because of fetal heart affection and hydrops. The lowest NT value in affected pregnancies was 1.0 mm for a T21-affected pregnancy and 1.3 mm for a T13-affected pregnancy, quite close to the suggested truncation limit of 0.8 mm in unaffected pregnancy. To prevent misleading results at extreme values, a low truncation limit for NT was set to 1.2 mm at a CRL of 45 mm, increasing to 1.8 mm at a CRL of 70–84 mm, and an upper truncation limit of 10 mm (16). Different truncation limits have been suggested for hCGB (MoM), 0.5–5.0 (17), 0.2–5.0 (18), and 0.3–5.0 (14), and for PAPP-A (MoM) 0.2–2.0 (17), 0.2–5.0 (18), and 0.2–3.0 (14). In this study, we have not applied truncation limits to NT (MoM), hCGB (MoM), PAPP-A (MoM), or LRs in our own programs, in either the univariate or the trivariate version. A set limit for the LR could be outside 1 in 2. Use of truncation limits in a software package can have considerable influence on the risk calculation (19), and the producer should provide information whether such limits are used, and if so, their values.

In conclusion, because our commercial software program for risk calculation (Astra) did not allow use of our own medians for NT, hCGB, and PAPP-A, we had to develop our own risk programs to find out whether the screening efficacy could be improved by using our own medians. At a given risk cutoff value, DRs for T13, T18, and T21 were slightly improved using our own program, based on trivariate normal distribution of NT (MoM), hCGB (MoM), and PAPP-A (MoM), but for a slightly higher FPR, especially in the very early weeks of the first trimester compared with the commercial program. The lower FPR in the commercial program was mainly due to an inaccuracy in the PAPP-A median. Although the default medians may work well in the startup phase, we should require from the software developers that locally estimated, center-specific medians and distribution parameters may be used to ensure high DRs and low FPRs for all 3 trisomies at a given risk cutoff, and that clear information is given about underlying risk algorithms and truncation limits.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References


