PRENATAL DIAGNOSIS
Prenat Diagn 2006; **26**: 730–739.
Published online 30 June 2006 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/pd.1497

Repeated measurement of pregnancy-associated plasma protein-A (PAPP-A) in Down syndrome screening: A validation study

Glenn E. Palomaki[1]*, David E. Wright[2], Anne M. Summers[3], Louis M. Neveux[1], Christian Meier[3], Andrea O'Donnell[3], Tianhua Huang[3], George J. Knight[1] and James E. Haddow[1]

[1] *Women & Infants Hospital, Providence, Rhode Island, USA*
[2] *University of Plymouth, Plymouth, United Kingdom, UK*
[3] *North York General Hospital, Toronto, Canada*

Objectives To confirm that measuring pregnancy-associated plasma protein-A (PAPP-A) in both first- and second-trimester serum samples improves Down syndrome screening.

Methods We selected paired first- and second-trimester stored serum samples from 34 Down syndrome pregnancies (cases) and 514 unaffected pregnancies (controls) and tested the second-trimester samples for PAPP-A and dimeric inhibin-A (DIA). First-trimester PAPP-A measurements were already available, as were second-trimester measurements of alpha-fetoprotein, unconjugated estriol (uE3), and human chorionic gonadotrophin (hCG).

Results PAPP-A was lower among cases than controls (0.47 MoM) in the first trimester (at an average of 12.5 weeks); in the second trimester, it was not different (0.91 MoM). Using repeated measures of PAPP-A alone, 21 of 34 cases were detected (62%, 95%CI 44% to 78%) with 5% false positives. At an observed 2% false-positive rate, the detection rates (DR) for the quadruple (69%) and serum integrated (69%) tests were lower than for the repeated measures test (75%). Modelled performance at 12 weeks was similar to these observed findings (70, 75, and 82%, respectively). If the first-trimester samples were collected at 10 weeks, however, DR would be higher (70, 81, and 91%, respectively).

Conclusions Adding a repeated measure of PAPP-A to existing serum markers improves Down syndrome screening to levels that are currently obtainable only by including ultrasound measurement of nuchal translucency (NT). Serum-based screening has the advantages of higher availability and reliability at a lower cost, resulting in a more effective screening strategy. A serum-based repeated measures test has a place in routine Down syndrome screening. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: Down syndrome; serum screening; repeated measures; PAPP-A

INTRODUCTION

Prenatal screening for Down syndrome is now offered routinely in the United States (Palomaki *et al*., 1997) and other countries, but the time in pregnancy when testing is done and the marker combinations vary widely (Mennuti and Driscoll, 2003; Wald *et al*., 2003; Wapner *et al*., 2003; Malone *et al*., 2005). Use of three or four second-trimester maternal serum measurements is commonplace, but women are increasingly being offered late first-trimester testing based on ultrasound and serum markers combined. In some areas, markers obtained in the first and second trimesters are being interpreted together as the integrated test (Knight *et al*., 2005), and variants such as sequential or contingent screening are being considered (Wright *et al*., 2004; Benn *et al*., 2005; Palomaki *et al*., 2006; Wright *et al*., 2006). Ultrasound measurement of nuchal translucency (NT) is the lynchpin of first-trimester screening, but there are barriers to its widespread introduction, which include special training requirements for sonographers, coupled with certification and ongoing quality assurance to ensure quality results (Malone, 2005). Costs of NT measurements in the United States are also high, compared to serum screening. Screening based on serum measurements alone is likely to remain widespread because of its price to performance ratio, availability, and reliability. Improving such testing, therefore, remains a priority.

A recent report (Wright and Bradbury, 2005) describes a new approach to identify useful serum markers that relies on measuring a given marker at another time during pregnancy, when the marker is less useful. In order for this repeated measurement to be useful, the second marker measurement (in MoM) needs to be highly correlated with the first. Preliminary modeling indicates that this strategy will provide substantial improvement to serum-based screening when the first-trimester sample is obtained at around 10 weeks' gestation. The current study examines the effect of repeated measures for the candidate marker, pregnancy-associated plasma protein-A (PAPP-A), using a new case/control dataset. A Gaussian risk algorithm that incorporates parameter estimates based on the literature and on other independent and

*Correspondence to: Glenn E. Palomaki, Division of Medical Screening, 70 Elm Street, 2nd Floor, Providence, RI 02903, USA. E-mail: gpalomaki@ipmms.org

Received: 4 April 2006
Revised: 4 April 2006
Accepted: 10 April 2006
Published online: 30 June 2006

original datasets is employed. The empirical results are compared with results obtained from modeling.

METHODS

Case/control sample set

A cohort of 19 176 women received integrated screening at North York General Hospital between December 1, 1999 and October 31, 2003. After testing, the first- and second-trimester maternal serum samples were stored at −20 °C. Demographic and pregnancy-related information, such as maternal age, gestational age, and maternal weight, were stored in the screening program's database (Alpha Logical Medical Systems, London, England). Ultrasound-based gestational age was between 10 and 13 completed weeks in the first trimester and 14 and 20 completed weeks in the second trimester. Pregnancy outcome, including whether the pregnancy was affected by Down syndrome, was available from the Ontario Multiple Marker Screening Database (Summers *et al*., 2003). Research Ethics Board (REB) approval was obtained for a case/control study of repeated measures screening for Down syndrome. For each pair of samples obtained from a documented singleton Down syndrome pregnancy (case), 10 paired sample sets from singleton pregnancies not known to be affected with any chromosomal abnormality were identified from the same period (within 1 month). Control samples (but not the case sample) from a case/control series of trisomy 18 pregnancies were also included.

Biochemical test results

First-trimester PAPP-A and second-trimester alpha-fetoprotein (AFP), unconjugated estriol (uE3), and human chorionic gonadotrophin (hCG) measurements in maternal serum (PerkinElmer Life and Analytical Sciences, Woodbridge, Ontario, Canada) were already available in the Ontario Multiple Marker Screening Database. All results had been converted to multiples of the median (MoM) and corrected for maternal weight. The second-trimester sera were thawed and tested for dimeric inhibin-A (DIA) (Diagnostic Systems Laboratories, Webster, TX) and PAPP-A (PerkinElmer). The first-trimester samples were tested for PAPP-A after a 1 : 5 dilution (according to package insert instructions). The matching second-trimester samples were tested in the same manner, but at a dilution of 1 : 40. The second-trimester DIA and PAPP-A measurements were converted to MoM, using medians derived from the control samples and were adjusted for maternal weight using existing equations. All measurements were made without knowledge of whether the sample was from a case or control pregnancy. Because a relatively large proportion of samples was from Asian women, a separate adjustment was used for existing markers to ensure that the median MoM was 1.0 in both Asian and non-Asian women. For PAPP-A and DIA measurements, a race-specific weight adjustment was performed prior to the racial factors being computed.

Statistical analysis

Each case and control pregnancy was assigned three Down syndrome risk estimates, on the basis of maternal age and various biochemical markers. In all instances, the risks were based on a model utilizing overlapping multivariate Gaussian distributions. Adding the second-trimester repeated PAPP-A measurements to serum integrated testing to form the repeated measure model involves expanding from five to six dimensions. This is similar to the way in which quadruple marker screening was expanded to serum integrated screening when first-trimester PAPP-A measurements were included. The first Down syndrome risk was based on the four second-trimester markers in combination with age—the quadruple test. The second also included PAPP-A measurements in the first trimester—the serum integrated test. The third included a repeated measure of PAPP-A in the second trimester—PAPP-A repeated measures test. All these three relied on the SURUSS parameter set (Wald *et al*., 2003), except for those parameters associated with the second-trimester PAPP-A measurements (logarithmic means and standard deviations in cases and controls, and correlation coefficients between the two PAPP-A measurements in cases and controls). A summary and re-analysis of published second-trimester PAPP-A measurements in case and control pregnancies were performed to create a final set of parameters used to verify this implementation of repeated measures. Standard deviations were determined by a linear regression of data from a probability plot between the 10th and 90th centiles. The age-associated risk was taken from a published equation (Hecht and Hook, 1996).

Correlation study sample set

A separate consecutive series of 838 women was also identified in order to study the within-woman correlation between first- and second-trimester PAPP-A MoM levels. These women received integrated screening during March and April, 2005, reported having a singleton pregnancy, and their maternal weight and race were available. All measurements were performed on fresh serum samples. PAPP-A measurements were then converted to MoM and adjusted for weight and race. The number of weeks between sampling was also computed.

Truncation limits

Historically, fixed truncation limits have been set for each Down syndrome marker independently. This has worked well because marker combinations used to date had pair-wise R-squared values near 0; nearly always less than 0.2. This method of truncating is still acceptable for quadruple and serum integrated risks. However, repeated PAPP-A measurements are highly

DOI: 10.1002/pd

correlated (R-squared values of 0.5 or higher). In order to compute reliable risks for repeated measures, we utilized a set of bivariate truncation limits for the first- and second-trimester PAPP-A measurements. Likelihood ratios were computed only for those values that lay within the 99% bivariate contour for both unaffected and Down syndrome pregnancies. PAPP-A paired values falling outside this area had their likelihood ratio set to the value found at the intersection of the contour's edge and a straight line between the center of that contour and the paired values. Standard univariate truncation limits were utilized for the remaining quadruple markers.

RESULTS

Correlation study

Figure 1 shows paired first- and second-trimester PAPP-A results for the correlation sample of 838 women. Five observations outside the 99.9% bivariate contour (shown by the ellipse) were excluded, leading to a trimmed correlation coefficient of 0.8146 (95%CI 0.7902 to 0.8361; 0.8044 before trimming). The relationship between the repeated measures and the time between samples was examined by computing trimmed correlation coefficients for paired samples with sampling intervals of 3.5 or fewer weeks, 3.5 to 4.5 weeks, and 4.5 weeks or more. The number of samples (after trimming), mean interval in weeks, and the trimmed correlation coefficients were 285, 2.9 weeks, 0.8652; 362, 4.0 weeks, 0.8078; and 186, 5.3 weeks, 0.7677, respectively, for the three sampling intervals. Although there was a trend towards higher correlations with shorter intervals, the use of a single correlation coefficient was considered acceptable over the range studied.

Description of the case/control study population

The average ages (standard deviation) of the case and control pregnancies were 33.9 (4.4) and 35.9 (3.6), respectively. In order to more closely simulate the general population, we subtracted an average of six years (random values drawn from a Gaussian distribution with mean 6 years, standard deviation 4 years) from all maternal ages. This resulted in both an average maternal age (27 years) and a distribution shape in the unaffected pregnancies that closely approximated that of the pregnancy population in the United States in 2000. This allows a direct comparison of observed and modelled risks and risk cut-off levels. The average first-trimester gestational age was 12.5 weeks, with 16%, 59%, and 25% at 11, 12, and 13 completed weeks, respectively. Asian women were an important subgroup in our study population, encompassing 28% of case pregnancies and 32% of control pregnancies. Nearly all the remaining pregnancies were Caucasian (3% Black). We accounted for differences in the median AFP, uE3, and hCG measurements among Asian women with unaffected pregnancies of 3%, 18%, and 16%, respectively. The average

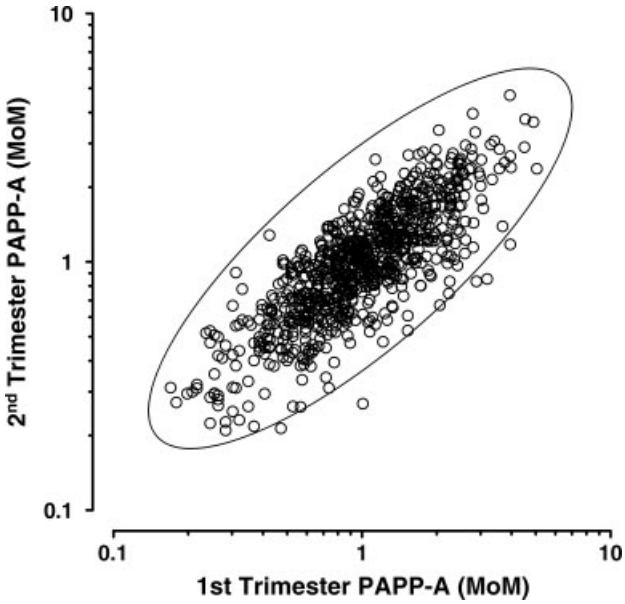

Figure 1—Scatterplot of paired first- and second-trimester pregnancy-associated plasma protein-A (PAPP-A) measurements in 838 women with unaffected pregnancies. These women had routine integrated Down syndrome screening at North York General Hospital. The horizontal logarithmic axis shows the first-trimester PAPP-A result in multiples of the median (MoM). That same woman's second-trimester PAPP-A MoM level is shown on the vertical logarithmic axis. The ellipse indicates the 99.9% bivariate contour and is used to identify five outlying values. The trimmed correlation coefficient is 0.8146

weight among Asian women was 127 pounds, compared to 143 pounds in the remaining control population. We then applied a race-specific maternal weight-adjustment equation. Median second-trimester PAPP-A and DIA measurements among Asian women differed by 31% and 13%, respectively. Subsequent adjustments were made such that the median analyte levels for PAPP-A and DIA were 1.00 in both Asian and non-Asian control pregnancies.

Repeated measures effect

Paired first- and second-trimester maternal serum samples and existing clinical information were available for 34 Down syndrome and 514 control pregnancies. Figure 2 shows first- and second-trimester measurements of PAPP-A for these pregnancies, expressed in MoM. First-trimester screening performance can be visualized by ignoring the vertical dimension. As expected, first-trimester measurements in the 34 Down syndrome cases are lower (median MoM 0.47, logarithmic standard deviation 0.2423) than in the 514 control pregnancies (1.00, 0.2458). Overall, 15 of the 34 cases (44%, 95% confidence interval 27–62%) fall at or below 0.41 MoM, along with 5% of controls. Second-trimester screening performance can be seen by ignoring the horizontal dimension. Viewed this way, second-trimester PAPP-A MoM levels in cases (median MoM 0.91, logarithmic standard deviation 0.2483) and controls (median MoM 1.00, logarithmic standard deviation 0.2449) are not significantly different ($t = 0.51$,

DOI: 10.1002/pd

$p = 0.61$). Observed correlation coefficients (after trimming at the 99.9th centile) are 0.7420 and 0.7567 for Down syndrome and unaffected pregnancies, respectively.

The combined performance of repeated PAPP-A measurements is demonstrated by the inclined line in Figure 2. This line is drawn at approximately the 95th centile of the bivariate distribution of unaffected pregnancies, and 21 of 34 Down syndrome samples fall on or above this line, indicating a detection rate of 62% (95% confidence interval 44–78%).

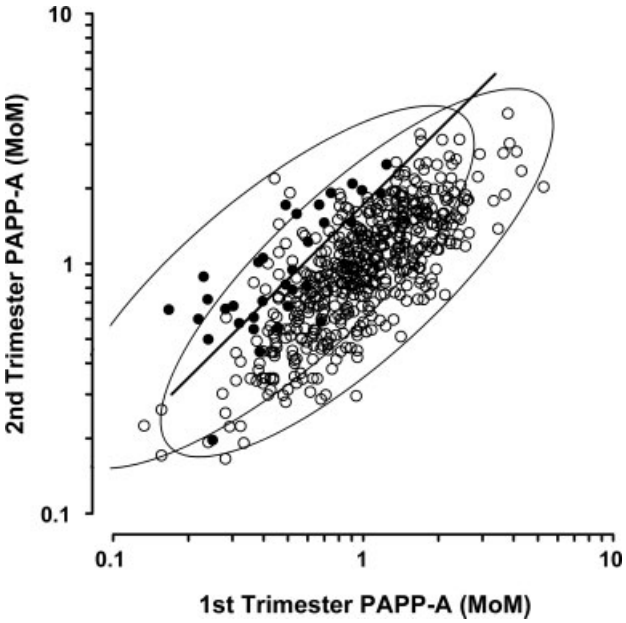

Figure 2—Scatterplot of paired first- and second-trimester pregnancy-associated plasma protein-A (PAPP-A) measurements in 34 women with a Down syndrome pregnancy (●) and 514 women with an unaffected pregnancy (○). The horizontal logarithmic axis shows the first-trimester PAPP-A result in multiples of the median (MoM). That same woman's second-trimester PAPP-A MoM level is shown on the vertical logarithmic axis. The solid line indicates the approximate 95th centile of the bivariate distribution for unaffected pregnancies. Overall, 62% of the Down syndrome observations (21 of 34) fall above this line. The two ellipses are drawn at the 99% bivariate contour for the two distributions

Identifying appropriate repeated measures parameters

To use our case/control dataset to validate an existing repeat measures algorithm, it is necessary to choose population parameters that are derived from other published data. The SURUSS report contains the only existing dataset with all necessary parameters (Wald *et al*., 2003). Most SURUSS parameters have been compared to other published parameter sets and then adjusted to take into account any important identified differences. This has not been systematically done, however, for the second-trimester PAPP-A measurements. We have performed this comparison and computed revised second-trimester PAPP-A parameters entirely based on the literature (Appendix).

Observed Down syndrome screening performance

Table 1 shows the observed screening performance of quadruple, serum integrated and PAPP-A repeated measures in 32 Down syndrome and 508 unaffected pregnancies (two cases and six controls had insufficient sample to measure second-trimester DIA and have been removed from the analysis). The first three rows report performance at fixed false-positive rates (FPRs), the next three rows at fixed detection rates (DR), and the last four rows at fixed risk cut-off levels. For example, at a fixed detection rate of 80%, the corresponding FPRs are 11%, 4.7%, and 2.6% for quadruple, serum integrated, and

Table 1—Observed false-positive rates (FPR), detection rates (DR), and second-trimester risk cut-off levels for three Down syndrome risk protocols among 32 Down syndrome and 508 unaffected pregnancies at 11 through 13 weeks' gestation

Constant	Quadruple[a]		Serum integrated[b]		PAPP-A repeated measures[c]	
FPR (%)	DR (%)	Cut-off[d]	DR (%)	Cut-off	DR (%)	Cut-off
1	59	1 : 70	59	1 : 60	66	1 : 50
2	69	1 : 100	69	1 : 95	75	1 : 120
3	72	1 : 150	75	1 : 170	81	1 : 170
DR (%)	FPR (%)	Cut-off	FPR (%)	Cut-off	FPR (%)	Cut-off
70	1.4	1 : 90	1.8	1 : 90	1.0	1 : 50
80	11	1 : 640	4.7	1 : 300	2.6	1 : 160
90	18	1 : 1400	6.7	1 : 510	4.9	1 : 330
Cut-off[d]	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)
1 : 100	69	2.0	69	2.0	75	1.4
1 : 150	72	3.0	75	2.8	78	2.2
1 : 200	72	3.9	75	3.3	81	3.3
1 : 250	72	4.9	75	4.1	84	3.7

[a] Quadruple test—maternal age, in combination with second-trimester measurements of maternal serum alpha-fetoprotein, unconjugated estriol, human chorionic gonadotrophin, and dimeric inhibin-A.
[b] Serum integrated test—the quadruple test and a first-trimester pregnancy-associated plasma protein-A (PAPP-A) measurement.
[c] PAPP-A repeated measures test—the serum integrated test and a second-trimester PAPP-A measurement.
[d] Second-trimester Down syndrome risk.

DOI: 10.1002/pd

repeated measures screening, respectively. At all cut-off levels shown in the table, the highest performance is associated with the PAPP-A repeated measure algorithm. In a few instances, the estimates appear counterintuitive. For example, at a risk cut-off level of 1 : 100, both quadruple and serum integrated screening have identical performance (69% detection at a 2.0% false-positive rate). One expects that serum integrated screening should show improved performance—and it does for most of the comparisons in the table. These discrepancies are likely due to the relatively small number of Down syndrome cases studied (32).

Figure 3 shows the Down syndrome risks generated by the three risk algorithms for the 32 cases and 508 control pregnancies. The figure includes the 1st, 2nd, 5th, and 10th centiles for unaffected pregnancies, along with the five unaffected pregnancies with the highest risks. Overall, the protocols that include more markers assign higher risks to the Down syndrome cases and lower risks to controls. For quadruple, serum integrated, and repeated measures strategies, the median risks (10th–90th centiles) are 1 : 40 (>1 : 1 to 1 : 1900); 1 : 19 (>1 : 1 to 1 : 750); and 1 : 18 (>1 : 1 to 1 : 390). The corresponding risks for unaffected pregnancies are 1 : 8900 (1 : 570 to 1 : 76 000); 1 : 9700 (1 : 800 to 1 : 90 000); and 1 : 36 000 (1 : 1100 to < 1 : 500 000).

Modelled Down syndrome screening performance

Table 2 contains the modelled, rather than the observed, screening performances for the same combinations of markers and parameters described earlier, using the maternal ages at delivery in the United States in 2000 as the baseline population. Results are provided at three FPRs, three DR, and four second-trimester risk cut-off levels (column 1). The next two columns provide screening performance for the quadruple test. Given that the performance is dependent on the gestational age of the first sample, we have chosen to present performance of the serum integrated test and the repeated measures test at two different first-trimester gestational ages. The middle four columns of results use the PAPP-A parameters at 10 weeks' gestation, along with a lower correlation coefficient (0.7677). These columns represent optimal conditions for both serum integrated and repeated measures Down syndrome screening (between 9 and 11 weeks' gestation). The right hand four columns of results (under the heading of 12 weeks' gestation) use the PAPP-A parameters at this less optimal time, for comparison with the observed results in Table 1.

When comparing either serum integrated or repeated measures testing, performance using the first-trimester samples collected at 10 weeks' gestation is better than with samples collected later in the first trimester (Table 2). For example, at a 1% false-positive rate, serum integrated screening with the first sample at 10 weeks is expected to detect 76% of Down syndrome cases, compared with 68% if the first sample is collected 2 weeks later. Similar improvements are seen for the repeated measures test with DR of 86% and 75%, respectively. The performance estimates for repeated measures at 10 weeks utilizes the lower correlation coefficient between repeated PAPP-A measurements of 0.7677 (observed for the 186 samples collected 4.5 or more weeks apart). As a way to estimate the impact of this parameter, we reran the modeling, changing only the correlation coefficient for first- *versus* second-trimester PAPP-A measurements in both Down syndrome and unaffected pregnancies to 0.8146 (the average correlation). At FPRs of 1–3%, use of the higher correlation

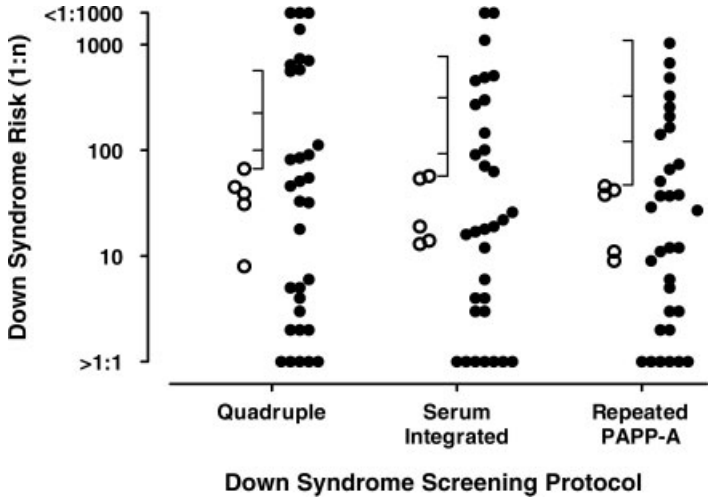

Figure 3—Down syndrome risks based on second-trimester quadruple testing, serum integrated testing, and serum integrated screening with repeated measures of pregnancy-associated plasma protein-A (PAPP-A) in 32 Down syndrome (●) and 508 unaffected pregnancies (○). The vertical logarithmic axis shows the second-trimester Down syndrome risk. The quadruple test combines maternal age with second-trimester measurements of alpha-fetoprotein (AFP), unconjugated estriol (uE3), human chorionic gonadotrophin (hCG) and dimeric inhibin-A (DIA). The serum integrated test includes the quadruple test and a first-trimester measurement of PAPP-A. The repeated measures test includes the serum integrated test and a second measurement of PAPP-A, this time in the second-trimester sample. The thin lines show the 1st, 2nd, 5th, and 10th centiles of unaffected pregnancies (from bottom to top). Among unaffected pregnancies, only the five with the highest risks (below the 1st centile) are shown individually, as open circles

DOI: 10.1002/pd

Table 2—Comparing modelled false-positive rates (FPR), detection rates (DR), and risk cut-off levels for three Down syndrome risk protocols for the United States pregnant population at 10 and 12 completed weeks' gestation

			At 10 completed weeks' gestation				At 12 completed weeks' gestation			
	Quadruple[a] testing		Serum integrated[b]		Repeated measures[c]		Serum integrated[b]		Repeated measures[c]	
FPR (%)	DR (%)	Cut-off[d]	DR (%)	Cut-off	DR (%)	Cut-off	DR (%)	Cut-off	DR (%)	Cut-off
1	62	1 : 45	76	1 : 65	86	1 : 80	68	1 : 40	75	1 : 45
2	70	1 : 75	81	1 : 130	91	1 : 180	75	1 : 80	82	1 : 90
3	75	1 : 110	85	1 : 200	93	1 : 310	79	1 : 130	86	1 : 150
DR (%)	FPR (%)	Cut-off	FPR (%)	Cut-off	FPR (%)	Cut-off	FPR (%)	Cut-off	FPR (%)	Cut-off
70	1.6	1 : 75	0.5	1 : 40	<0.5	1 : 11	1.2	1 : 55	0.6	1 : 30
80	3.9	1 : 200	1.6	1 : 100	<0.5	1 : 30	3.4	1 : 130	1.3	1 : 75
90	10	1 : 730	5.9	1 : 450	1.1	1 : 90	9.9	1 : 440	4.9	1 : 230
Cut-off[d]	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)
1 : 100	75	2.8	81	2.1	89	1.5	77	2.4	83	2.1
1 : 150	79	4.0	84	2.9	91	2.1	81	3.5	86	2.9
1 : 200	82	5.2	86	3.8	91	2.6	83	4.5	88	3.7
1 : 250	84	6.3	87	4.6	93	3.0	85	5.5	89	4.4

[a] Quadruple test—maternal age, in combination with second-trimester measurements of maternal serum alpha-fetoprotein, unconjugated estriol, human chorionic gonadotrophin, and dimeric inhibin-A.
[b] Serum integrated test—the quadruple test and a first-trimester pregnancy-associated plasma protein-A (PAPP-A) measurement.
[c] PAPP-A repeated measures test—the serum integrated test and a second-trimester PAPP-A measurement.
[d] Second-trimester Down syndrome risk.

increases detection by 2 to 3% points. At DR of 70% and 80%, FPRs are still less than 0.5%. At 90% detection, the false-positive rate improves from 1.7–1.1%. At the four risk cut-off levels, detection increases slightly (1 or 2% points) while FPRs decrease slightly (by 0.2–0.5% points). These analyses indicate that performance is dependent on the correlation coefficient, but that there were no major alterations to the detection and FPRs found over the range examined.

For purposes of comparison, Table 3 provides performance estimates for three Down syndrome screening strategies: PAPP-A repeated measures screening with the first sample collected at 10 weeks' gestation (as shown in Table 2), full integrated screening (serum integrated screening with a first-trimester NT measurement) at 12 weeks' gestation, and first-trimester combined screening (NT, PAPP-A, and free beta-hCG measurements), again at 12 weeks' gestation. The performance at 12 weeks' gestation has been chosen for the combinations that include NT measurements, as those measurements are recommended to be performed between 11 and 13 completed weeks' gestation. PAPP-A repeated

Table 3—Comparing modelled false-positive rates (FPR), detection rates (DR), and risk cut-off levels for three Down syndrome screening strategies

	Repeated measures testing at 10 weeks' gestation[a]		Integrated testing at 12 weeks' gestation[b]		Combined testing at 12 weeks' gestation[c]	
FPR (%)	DR (%)	Cut-off[d]	DR (%)	Cut-off	DR (%)	Cut-off
1	86	1 : 80	82	1 : 70	69	1 : 50
2	91	1 : 180	87	1 : 150	75	1 : 100
3	93	1 : 310	90	1 : 230	79	1 : 150
DR (%)	FPR (%)	Cut-off	FPR (%)	Cut-off	FPR (%)	Cut-off
70	<0.5	1 : 11	<0.5	1 : 16	1.1	1 : 60
80	<0.5	1 : 30	0.7	1 : 55	3.3	1 : 170
90	1.1	1 : 90	3.2	1 : 250	11.3	1 : 630
Cut-off[d]	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)
1 : 100	89	1.5	85	1.5	75	2.0
1 : 150	91	2.1	87	2.1	79	3.0
1 : 200	91	2.6	89	2.7	81	3.9
1 : 250	93	3.0	90	3.3	83	4.8

[a] Maternal age and a first-trimester pregnancy-associated plasma protein-A (PAPP-A) with second-trimester measurements of alpha-fetoprotein (AFP), unconjugated estriol (uE3), human chorionic gonadotrophin (hCG), dimeric inhibin-A (DIA), and PAPP-A.
[b] Maternal age and first-trimester nuchal translucency and PAPP-A measurements with second-trimester measurements of AFP, uE3, hCG, and DIA.
[c] Maternal age in combination with first-trimester nuchal translucency, PAPP-A, and free beta-hCG measurements.
[d] Second-trimester Down syndrome risk.

DOI: 10.1002/pd

measures screening is somewhat better than integrated screening. At a 2% false-positive rate, repeated measures screening is expected to detect about 91% of Down syndrome pregnancies, while integrated screening detects 87%. Combined screening has the lowest Down syndrome screening performance. At the same 2% false-positive rate, the detection rate is 75%. Replacing free beta-hCG measurements with total (or intact) hCG measurements resulted in the same detection rate of 75% (other estimates in Table 3 were also equivalent with substitution).

DISCUSSION

When measured on two occasions, PAPP-A acts as a strong bivariate marker for Down syndrome. In the first trimester, PAPP-A MoM values are much lower in Down syndrome pregnancies than in unaffected pregnancies. In contrast, second-trimester MoM values for PAPP-A in these two groups are essentially the same. Importantly, within the same woman, the two PAPP-A MoM values are highly correlated (e.g. a 'low' measurement in the first trimester is expected to be followed by a 'low' measurement in the second trimester). Although counterintuitive, adding the second trimester PAPP-A measurement improves performance because it acts as the woman's own reference value. This allows for a more informed interpretation of the first-trimester result. Figure 4 provides a hypothetical example of a woman whose first-trimester PAPP-A value is 0.7 MoM. The 0.7 MoM value is at the 25th centile for unaffected pregnancies, but is at the 70th centile for Down syndrome pregnancies. The high correlation means that the following can be expected: (1) if the pregnancy is affected with Down syndrome, that woman's second-trimester PAPP-A value will shift to 1.4 MoM (about the 70th centile), and (2) if the pregnancy is unaffected, the woman's second-trimester PAPP-A value will remain near 0.7 MoM (about the 25th centile).

If the first-trimester PAPP-A value of 0.7 MoM were to be interpreted alone, the likelihood ratio would be about 1.0 (equivalent to the height from baseline to the dashed 'Down syndrome' curve divided by the height from baseline to the solid 'unaffected' curve). The likelihood ratio associated with the repeated PAPP-A measurements in the Down syndrome pregnancy (0.7 MoM, 1.4 MoM) would shift from 1.0 to 8.0, while the repeated measure in the unaffected pregnancy (0.7 MoM, 0.7 MoM) would shift from 1.0 to 0.06. These widely separated likelihood ratios are the reason for the improved performance.

The original report of repeated measures may have overestimated the impact of repeated PAPP-A measurements, due to problems with the reported second-trimester PAPP-A parameters and the use of univariate truncation limits. We have addressed both of these, and the modelled performance of repeated PAPP-A measurements (Table 2) is likely to be similar to that found in an everyday setting. Our observed performance at an average of 12 completed weeks' gestation is consistent with the modelled results at that time in gestation. However, there is one *caveat*. Modeling, by its nature, considers only Down syndrome and unaffected pregnancies. Other normal and abnormal pregnancy outcomes are not considered. In Figure 1, for example, only one of the 838 observations might be expected outside the 99.9% bivariate contour. However, five were observed. When modeling screening performance with FPRs of 5% or higher, these infrequent observations are less relevant. However, repeated measures protocols are likely to have targeted FPRs of 2%, 1%, or even lower. At these low rates, infrequent occurrences will have a larger impact, and the modelled FPRs might be somewhat less reliable.

Our study did not address other repeated measures. Prime candidates include uE3 and hCG in the first trimester. As with PAPP-A, earlier first-trimester measurements (e.g. 10 weeks' gestation) are likely to

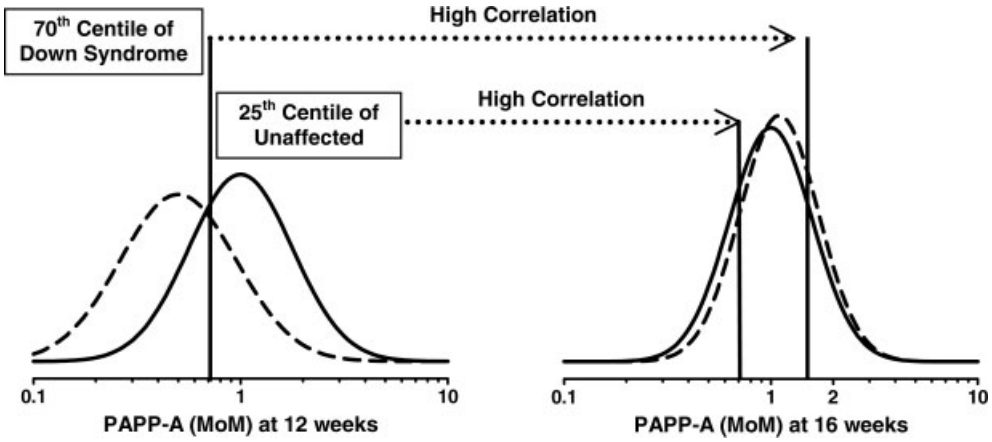

Figure 4—A hypothetical example showing how the second-trimester PAPP-A test result can improve the interpretation of the first-trimester PAPP-A test result. A woman's first-trimester serum PAPP-A result is 0.7 MoM; about the 70th centile of the Down syndrome curve, but only the 25th centile of unaffected pregnancies. Because of the high within-woman, between-trimester PAPP-A correlations for both Down syndrome and unaffected pregnancies, a repeated PAPP-A result near the 25th centile (consistent with unaffected pregnancies) would reduce the risk, while one near the 70th centile would increase the risk (consistent with Down syndrome pregnancies)

DOI: 10.1002/pd

perform better than those obtained later (e.g. 12 or 13 weeks' gestation). However, assay performance must first be verified at these gestational ages. For example, the levels of uE3 are measurable at 11 to 13 weeks' gestation (SURUSS), but it is not clear that this would be possible at 10 weeks' and earlier. Our experience with serum integrated screening, offered throughout Maine, has shown that many pregnant women will be able to provide a serum sample early in pregnancy. More than half the 11 159 women who provided a first-trimester sample did so at 10 weeks' gestation or earlier (Knight *et al*., 2005). A repeated measures strategy might need to be uncoupled from NT measurements, which are most often collected at around 12 weeks' gestation.

The PAPP-A repeated measures algorithm is associated with higher performance than serum integrated testing or quadruple testing. Taken together, our observed and modelled estimates for repeated measures indicate that it may be possible to reach the performance of a fully integrated test (serum integrated with an NT measurement) by measuring serum markers alone if the first-trimester sample is targeted to be collected between 9 and 11 weeks' gestation. Repeated measures testing has three major advantages over protocols that include NT measurements: availability, reliability, and affordability. Serum-based screening is already available to over 2 million women who opt for second-trimester serum screening annually in the United States. Extensive internal and external quality assessment schemes that have both state and federal oversight are already in place for prenatal screening laboratories. Adding one (or two) serum tests is more affordable than a separate ultrasound study, especially if NT measurements are done in the setting of a comprehensive first-trimester ultrasound scan. One major criticism of integrated screening with NT measurements is that of 'holding' information until the second trimester (Copel and Bahado-Singh, 1999; Cuckle, 2002; Wapner *et al*., 2003). Repeated measures testing avoids this problem because the first-trimester PAPP-A result by itself is of little practical value. During our serum integrated study in Maine (Palomaki *et al*., 2005), neither women nor their providers asked for interim results. A survey conducted as part of that study found that holding first-trimester results was acceptable to the women; only 7% of those surveyed indicated that it was difficult to wait until the second trimester to obtain their results. If a matching second-trimester sample is not obtained by a reasonable time (e.g. 18 weeks' gestation), either no risk is reported or a risk based only on the first-trimester biochemistry can be reported. Programs can also decide whether to test for PAPP-A in the fresh samples or hold the first-trimester sample and test only when the second-trimester sample arrives. However, before repeated measures testing is introduced into routine clinical practice, it should be subjected to formal pilot trials to document performance and explore issues relating to widespread implementation.

ACKNOWLEDGEMENTS

We thank Diagnostic Systems Laboratories (Webster, TX) and PerkinElmer Life and Analytical Sciences (Woodbridge, Ontario, Canada) for supplying reagents. The University of Plymouth (Dr Wright) holds a patent application for repeated measures.

APPENDIX: REVIEW AND ANALYSIS OF SECOND-TRIMESTER PAPP-A PARAMETERS FROM SURUSS AND OTHER PUBLISHED STUDIES

Table 4 provides a summary of second-trimester PAPP-A measurements in Down syndrome and unaffected pregnancies that includes those from SURUSS

Table 4—Published population parameters for second-trimester pregnancy-associated plasma protein-A (PAPP-A) measurements in Down syndrome and unaffected pregnancies

		Down syndrome pregnancies			Unaffected pregnancies			Median
Reference	PAPP-A method	N	Median (a)	SD	N	Median (b)	SD	Ratio (a/b)
(Cuckle *et al*., 1992)	Dako Ab[a]	18	0.87	0.155	90	0.99	0.234	0.88
(Wald *et al*., 1992)	Dako Ab	16	1.02	0.203	147	1.00	NR	1.02
(Knight *et al*., 1993)	Dako Ab	30	1.01	0.226	147	1.00	NR	1.01
(Spencer *et al*., 1994)	Dako Ab	12	1.08	0.230	NR	1.00	NR	1.08
(Berry *et al*., 1997)	Dako Ab	37	0.94	0.252	267	0.98	0.211	0.96
(Spencer *et al*., 2002)	Kryptor & PE Delfia	172	0.85	0.282	726	0.97	0.236	0.88
(Wald *et al*., 2003)[b]	PE Delfia (1 : 5/1 : 20)	75	1.11/1.18	0.187/0.220	375	1.00	0.218/0.255	1.11/1.18
Current study	PE Delfia (1 : 40)	NR	NR	NR	873	1.00	0.245	NR
Total/Average[c]		360		0.252	2685		0.240	0.97
Trimmed SD[d]		342		0.256	2358		0.243	

N, number of observations; SD, standard deviation; NR, not reported.
[a] The PAPP-A antibody available from Dako A/S (Glostrup, Denmark) was used according to published protocols.
[b] Revised medians and SDs are from Wald NJ *et al*., J Med Screen, 2006 (in press) after 1 : 20 dilution of samples over 10 000 IU/mL.
[c] The total number of samples in the column, or the average value of the column, weighted by the number of samples.
[d] Same as the average/total, except one study (Cuckle *et al*., 1992), which was excluded from estimating the Down syndrome SD, and one study (Berry *et al*., 1997), which was excluded from estimating the unaffected SD.

DOI: 10.1002/pd

(Wald *et al*., 2003) and other studies (Cuckle *et al*., 1992; Knight *et al*., 1993; Spencer *et al*., 1994; Berry *et al*., 1997) Wald and his colleagues recently revised SURUSS estimates for second-trimester PAPP-A parameters (Wald *et al*., 2006), by correcting the systematic underestimation of high PAPP-A values that resulted from inadequate sample dilution (second-trimester PAPP-A levels are about 10 times higher than first-trimester levels). The revised SURUSS parameters are contained in Table 4 and are more consistent with the existing literature.

Figure 5 shows a probability plot of second-trimester PAPP-A measurements in 79 Down syndrome pregnancies compiled from three studies reported in Table 4, which provided individual PAPP-A measurements (in MoM) as part of a figure or table (Knight *et al*., 1993; Spencer *et al*., 1994; Berry *et al*., 1997). Although the data are from three separate studies, an analysis of variance indicates that neither the mean levels nor the variances differ significantly (F-test for equality of means = 0.15, $p = 0.8$, Levene's test for homogeneity of variance $F = 0.99$, $p = 0.4$). The data, therefore, can be combined. Figure 6 shows a probability plot of second-trimester PAPP-A measurements in the 828 second-trimester samples (after trimming) from the North York General Hospital screening service. These data also fit a log Gaussian distribution (parameters provided in Table 4). On the basis of these figures, first-trimester PAPP-A measures in both unaffected and Down syndrome pregnancies fit a log Gaussian distribution well. The revised SURUSS probability plot (Wald *et al*., 2006) still deviates somewhat from the expected straight line at higher values, in contrast to the probability plots in Figures 5 and 6.

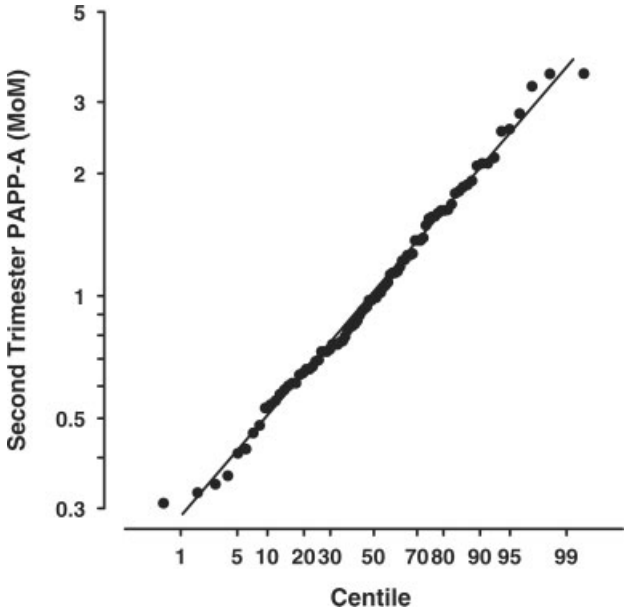

Figure 5—A probability plot of second-trimester pregnancy-associated plasma protein-A (PAPP-A) in 79 Down syndrome pregnancies from the literature. These observations were obtained from three studies. The papp-A MoM values for the 79 Down syndrome pregnancies are plotted vertically on a logarithmic scale and horizontally on the expected Gaussian centile—the latter based on the rank of the observation. When the points plotted follow an approximate straight line, the distribution is considered to be logarithmic Gaussian

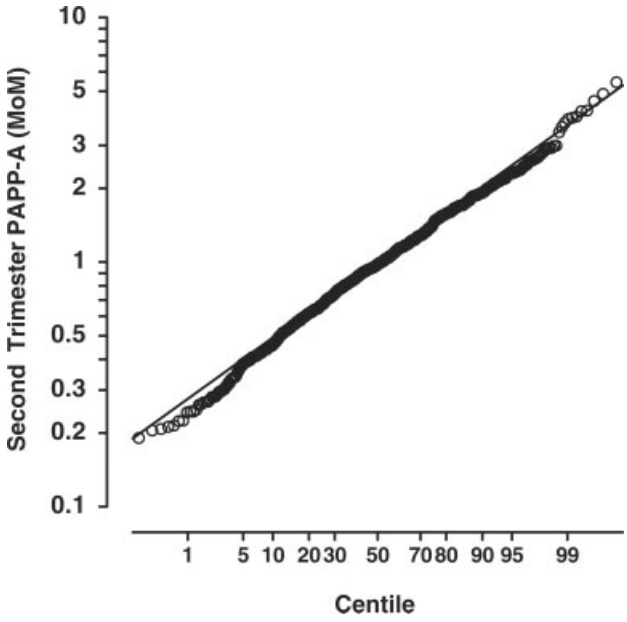

Figure 6—A probability plot of second-trimester pregnancy-associated plasma protein-A (PAPP-A) in 838 unaffected pregnancies. These observations are from the same women shown in Figure 1 and were obtained by testing stored samples from women having routine integrated Down syndrome screening. The PAPP-A MoM values for each of the unaffected pregnancies are plotted vertically on a logarithmic scale and horizontally on the expected Gaussian centile—the latter based on the rank of the observation. When the points plotted follow an approximate straight line, the distribution is considered to be logarithmic Gaussian

We use the summary parameters describing the second-trimester PAPP-A results in Down syndrome and unaffected pregnancies as shown in Table 4 for assigning risks and modeling the screening performance. We also use the correlation of 0.8146 from our 838 unaffected pregnancies (Figure 1). Besides our current case/control set, the only direct estimate of the correlation coefficient in Down syndrome cases is from SURUSS (0.6980). We use the same correlation coefficient of 0.8146 that was derived from our unaffected pregnancies as the correlation coefficient for Down syndrome pregnancies. This is done to maintain consistency of using literature-based estimates; the corrected SURUSS correlation (0.8262) for Down syndrome pregnancies is similar. In the subsequent modeling, all other parameters for AFP, uE3, hCG, DIA in the second trimester and PAPP-A in the first trimester, as well as truncation limits, are those originally published in the SURUSS reports. These values have been extensively compared to the literature and can be considered representative of all parameter sets.

REFERENCES

Benn P, Wright D, Cuckle H. 2005. Practical strategies in contingent sequential screening for Down syndrome. *Prenat Diagn* **25**: 645–652.

Berry E, Aitken DA, Crossley JA, Macri JN, Connor JM. 1997. Screening for Down's syndrome: changes in marker levels and

detection rates between first and second trimesters. *Br J Obstet Gynaecol* **104**: 811–817.

Copel JA, Bahado-Singh RO. 1999. Prenatal screening for Down's syndrome–a search for the family's values. *N Engl J Med* **341**: 521–522.

Cuckle HS. 2002. Growing complexity in the choice of Down's syndrome screening policy. *Ultrasound Obstet Gynecol* **19**: 323–326.

Cuckle H, Lilford RJ, Teisner B, *et al*. 1992. Pregnancy associated plasma protein a in Down's syndrome. *BMJ* **305**: 425.

Hecht CA, Hook EB. 1996. Rates of Down syndrome at livebirth by one-year maternal age intervals in studies with apparent close to complete ascertainment in populations of European origin: a proposed revised rate schedule for use in genetic and prenatal screening. *Am J Med Genet* **62**: 376–385.

Knight GJ, Palomaki GE, Haddow JE, *et al*. 1993. Pregnancy associated plasma protein a as a marker for Down syndrome in the second trimester of pregnancy. *Prenat Diagn* **13**: 222–223.

Knight GJ, Palomaki GE, Neveux LM, *et al*. 2005. Integrated serum screening for Down syndrome in primary obstetric practice. *Prenat Diagn* **25**: 1162–1167.

Malone FD. 2005. Nuchal translucency-based Down syndrome screening: barriers to implementation. *Semin Perinatol* **29**: 272–276.

Malone FD, Canick JA, Ball RH, *et al*. 2005. First-trimester or second-trimester screening, or both, for Down's syndrome. *N Engl J Med* **353**: 2001–2011.

Mennuti MT, Driscoll DA. 2003. Screening for Down's syndrome–too many choices? *N Engl J Med* **349**: 1471–1473.

Palomaki GE, Steinort K, Knight GJ, Haddow JE. 2006. Comparing three screening strategies for combining first- and second-trimester Down syndrome markers. *Obstet Gynecol* **107**: 367–375.

Palomaki GE, Knight GJ, McCarthy JE, Haddow JE, Donhowe JM. 1997. Maternal serum screening for Down syndrome in the United States: a 1995 survey. *Am J Obstet Gynecol* **176**: 1046–1051.

Palomaki GE, Kloza EM, Haddow JE, Williams J, Knight GJ. 2005. Patient and health professional acceptance of integrated serum screening for Down syndrome. *Semin Perinatol* **29**: 247–251.

Spencer K, Aitken DA, Crossley JA, *et al*. 1994. First trimester biochemical screening for trisomy 21: the role of free beta hCG, alpha fetoprotein and pregnancy associated plasma protein A. *Ann Clin Biochem* **31**: 447–454.

Spencer K, Crossley JA, Aitken DA, *et al*. 2002. Temporal changes in maternal serum biochemical markers of trisomy 21 across the first and second trimester of pregnancy. *Ann Clin Biochem* **39**: 567–576.

Summers AM, Farrell SA, Huang T, Meier C, Wyatt PR. 2003. Maternal serum screening in Ontario using the triple marker test. *J Med Screen* **10**: 107–111.

Wald NJ, Rodeck C, Hackshaw AK, *et al*. 2003. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *J Med Screen* **10**: 56–104.

Wald NJ, Rodeck C, Hackshaw AK, *et al*. 2006. *J Med Screen*, in press.

Wald N, Stone R, Cuckle HS, *et al*. 1992. First trimester concentrations of pregnancy associated plasma protein A and placental protein 14 in Down's syndrome. *BMJ* **305**: 28.

Wapner R, Thom E, Simpson JL, *et al*. 2003. First-trimester screening for trisomies 21 and 18. *N Engl J Med* **349**: 1405–1413.

Wright D, Bradbury I, Cuckle H, *et al*. 2006. Three stage contingent screening for Down syndrome. *Prenat Diagn* **26**: 528–534.

Wright DE, Bradbury I. 2005. Repeated measures screening for Down's Syndrome. *BJOG* **112**: 80–83.

Wright D, Bradbury I, Benn P, Cuckle H, Ritchie K. 2004. Contingent screening for Down syndrome is an efficient alternative to non-disclosure sequential screening. *Prenat Diagn* **24**: 762–766.

DOI: 10.1002/pd