Chapter 6. Combining first trimester markers

6.1 Introduction

First trimester testing for Down syndrome and other autosomal trisomies is often portrayed as being better than second trimester testing because of the potential for earlier diagnostic testing and, if an abnormality is identified, earlier termination of pregnancy. This would be true, if access to the first trimester diagnostic test (chorion villus sampling or CVS) were as available, safe and reliable as second trimester amniocentesis. In the U.S, CVS is less available, is generally considered to be associated with a higher procedure-related fetal loss, and is less reliable than amniocentesis. Given this, the performance of first trimester testing would need at least similar performance compared to the corresponding testing in the second trimester. For this reason, neither NT testing alone, nor biochemistry alone, would be considered sufficiently powerful to be offered as a routine first trimester test. The following sections will provide estimates of performance for various combinations of markers, but only those with both NT and biochemistry would be considered suitable for introduction as a first line test for Down syndrome; the major focus of such testing.

6.2 Modeling test performance

Chapter four provided the parameters necessary to describe the performance of serum screening markers for trisomy 18, with the best performers being PAPP-A, free β and intact hCG measurements. Chapter five provided similar information for ultrasound markers, with only NT measurements being suitable for routine use. Based on a multivariate overlapping Gaussian model, a *monte carlo* simulation was performed as a way to estimate screening performance of a combined (serum + ultrasound) model for identifying trisomy 18. The age-associated risk at term, and the proportion of affected fetuses spontaneously lost are taken from Chapter 2. Table 6.2-1 examines four tests, all of which use maternal age and PAPP-A measurements. Two are serum only, with the addition of either free β or hCG, while the latter two add NT measurements to these serum markers. Clinical testing would be expected to concentrate only on the last two columns, as these are needed for Down syndrome screening. The top half of the table shows false positive rates at fixed detection rates, while the bottom shows detection rates at fixed false positive rates.

Table 6.2-1. Modeled trisomy 18 detection rates (DR) and false positive rates (FPR) using first trimester maternal serum markers with and without ultrasound measurements of nuchal translucency (NT)

	Maternal age in combination with			
	PAPP-A & free β	**PAPP-A & hCG**	**PAPP-A, free β & NT**	**PAPP-A, hCG & NT**
DR (%)				
50	0.2	0.4	<0.1	<0.1
60	0.4	0.7	<0.1	<0.1
70	0.6	1.0	<0.1	<0.1
80	1.4	2.4	0.1	0.1
90	4.2	6.3	0.5	0.8
FPR (%)				
0.3	53	46	87	85
0.5	63	52	90	87
0.7	73	60	91	89
1.0	77	70	93	91
1.5	81	74	95	93

Testing using only serum markers is somewhat better with the use of free β than with hCG measurements. For example, at a detection rate of 80%, the combination of maternal age, PAPP-A and free β would require a 1.4% false positive rate. If hCG were to replace free β measurements, the false positive rate would increase to 2.4%. However, when NT measurements are also included, a false positive rate of only about 0.1% would be required for both combinations of serum markers.

Table 6.2-2 shows the results of modeling stratified by the trisomy 18 risk cut-off level and included an estimate of the odds of being affected given a positive test result (OAPR) in the first trimester. The risk cut-off levels are shown both at term, and in the second trimester. Term results allow comparison of test performance with that found for second trimester markers (Chapter 3), while the first trimester risks are probably more appropriate for clinical use. The combined test performance at a term risk of 1:100 (first trimester risk of about 1:30) yields about an 85% detection rate for a 0.2% to 0.3% false positive rate, regardless of which hCG marker is used. The corresponding OAPRs are in the order of 1:2 to 1:4. If higher performance was needed, a term risk cut-off of 1:200 would improve detection to 90%, but with a doubling of the false positive rate to about 0.5%. The corresponding OAPR would be about 1:6.

As shown in second trimester demonstration studies (Chapter 3.11), trisomy 18 test performance is extremely good, with high detection rates at very low false positive rates, usually below 0.3% (3 per 1000 false positives). At such low rates, the model will not agree well with rates found in practice. This is because other relatively rare outcomes (*e.g.,* fetal death, anencephaly) that the trisomy 18 algorithms preferentially identify can result in a noticeable increase in the false positive rate (*i.e.,* from 0.1% to 0.3%). However, most of these additional 'false positives' are not unaffected pregnancies.

Table 6.2-2. Modeled first trimester trisomy 18 screening performance at selected trisomy 18 risk cut-off levels

Risk (1:n) at term (in first[1])	Maternal age in combination with PAPP-A & free β			PAPP-A & hCG			PAPP-A, free β & NT			PAPP-A, hCG & NT		
	DR	FPR	OAPR[2]	DR	FPR	OAPR[2]	DR	FPR	OAPR[2]	DR	FPR	OAPR[2]
1: 20 (1: 6)	36	<0.1	>1: 3	29	0.1	1: 4	71	<0.1	>1: 2	66	<0.1	>1: 2
1: 50 (1:14)	56	0.4	1: 7	47	0.3	1: 7	79	0.1	1: 1	77	0.1	1: 1
1: 70 (1:20)	70	0.6	1: 9	56	0.5	1: 9	83	0.1	1: 1	80	0.2	1: 3
1:100 (1:29)	74	0.8	1:11	69	0.9	1:14	86	0.2	1: 2	83	0.3	1: 4
1:120 (1:35)	77	1.1	1:15	71	1.1	1:16	87	0.3	1: 4	85	0.3	1: 4
1:150 (1:43)	80	1.4	1:18	75	1.4	1:19	88	0.3	1; 4	86	0.4	1: 5
1:200 (1:60)	83	2.0	1:25	78	2.0	1:27	90	0.4	1: 5	88	0.6	1: 7

[1] assumes 72% fetal loss from 11-13 weeks to term (Table 2.4-1)

[2] with a first trimester prevalence of 9.61/10,000 (term = 2.69/10,000)

DR = detection rate, FPR = false positive rate, OAPR = first trimester odds of being affected given a positive result

The shaded rows indicate two potential cut-off levels that could be used for the first trimester combined test for trisomy 18

6.3 Summary of demonstration studies

Literature search: The literature was search for demonstration studies containing a prospective implementation of a trisomy 18 (or trisomy 18/13) specific protocol in the late first trimester. Most reports of trisomy 18 testing are contained within a report on Down syndrome screening performance. In addition to reporting the results of a prospective application of a trisomy 18 specific protocol using biochemistry and NT measurements in the late first trimester, it was necessary for the publication to provide the needed information to estimate both the detection rate and the false positive rate (no study was excluded because they did not report the detection rate). Studies that combined trisomy 18 and 13 together in either the algorithm or in computing the detection rate were included. The identification of demonstration studies was performed using the results of the NT literature search (Section 5.2). A total of 33 publications were individually reviewed and eight were found to be suitable for analysis (Orlandi *et al.*, 1997; Krantz *et al.*, 2000; Spencer *et al.*, 2000c; Wapner *et al.*, 2003; Stenhouse *et al.*, 2004; Breathnach *et al.*, 2007; Jaques *et al.*, 2007; Schaelike *et al.*, 2009). All used a combination of maternal age, PAPP-A, free β and NT measurements in the late first trimester. There were 25 excluded studies (Nicolaides *et al.*, 1994; Scott *et al.*, 1996; Zimmermann *et al.*, 1996; Taipale *et al.*, 1997; Theodoropoulos *et al.*, 1998; O'Callaghan *et al.*, 2000; Acacio *et al.*, 2001; Brizot *et al.*, 2001; Niemimaa *et al.*, 2001; Schuchter *et al.*, 2001; Tsai *et al.*, 2001; Wayda *et al.*, 2001; Bindra *et al.*, 2002; von Kaisenberg *et al.*, 2002; Chasen *et al.*, 2003; Cheng *et al.*, 2003; Spencer *et al.*, 2003; Borrell *et al.*, 2004; Nicolaides *et al.*, 2005; Wojdemann *et al.*, 2005; Dhaifalah *et al.*, 2006; Kim *et al.*, 2006; Perni *et al.*, 2006; Leung *et al.*, 2007; Torring, 2009). The main reasons for excluding these studies were: no specific trisomy 18 algorithm, no biochemistry, and a retrospective application of an algorithm that was developed using the same dataset (over-fitting). Because elevated NT measurements are a strong predictor of both trisomy 18 and Down syndrome, some programs did not think it was necessary to have a test interpretation focused on trisomy 18/13. Figure 6.3-1 shows the date of publication versus the number of trisomy 18 pregnancies studied.

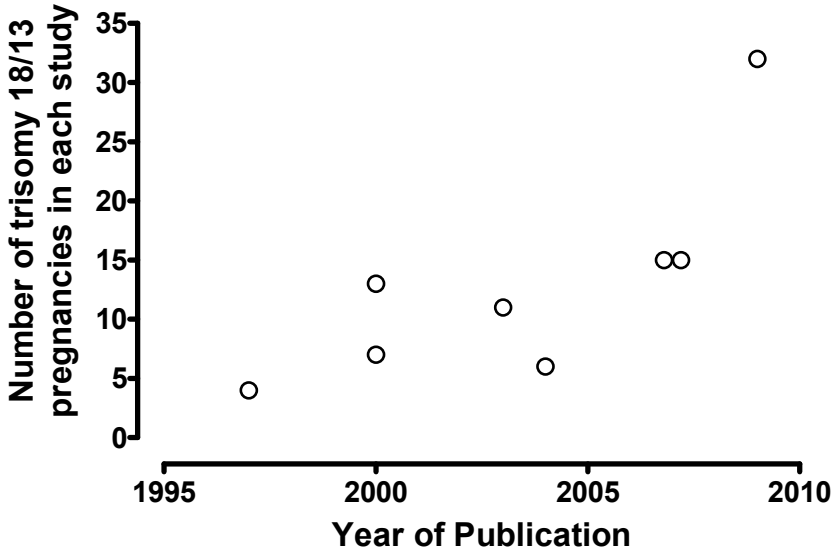

Figure 6.3.1 Summary of publications included in the analysis of demonstration studies for first trimester combined testing for trisomy 18.

Results: Table 6.3-1 shows relevant information about these eight publications, including the observed false positive rates and corresponding detection rates. The table is sorted in order of increasing false positive rates. Overall, 103 trisomy 18 pregnancies were identified in the eight studies, among over 87,000 women tested. One study (Schaelike *et al.*, 2009) included trisomy 13 and trisomy 18 together. In a general population of first trimester pregnancies, one would expect the trisomy 18 prevalence to be approximately 9.6/10,000, or about 1:1000 (see Table 6.2-2 footnote for the derivation). The overall rate in the eight studies was about 1:690. However, four studies appeared to have included referral patients many more older women, or both, as their rates were between 1:300 and 1:400 (Orlandi *et al.*, 1997; Krantz *et al.*, 2000; Spencer *et al.*, 2000c; Schaelike *et al.*, 2009). No single summary estimate for the detection rate and false positive rate are provided, as there is a clear relationship between increasing false positive rates and increasing detection rates.

Table 6.3-1. Results of eight demonstration studies of combined testing for trisomy 18.

Author	False Positive	True Negative	False Positive Rate (%)	True Positive	False Negative	Detection Rate (%)	Prevalence (1:n)
(Breathnach *et al.*, 2007)	116	35,793	0.32	19	30	63.3	1:1201
(Jaques *et al.*, 2007)	56	15,947	0.35	10	15	66.7	1:1071
(Schaelike *et al.*, 2009)	74	10,485	0.70	21	32[1]	65.6	1: 332
(Krantz *et al.*, 2000)	44	5,675	0.77	16	16	100	1: 360
(Stenhouse *et al.*, 2004)	53	4,921	1.07	7	8	87.5	1: 628
(Spencer *et al.*, 2000c)	45	3,717	1.20	8	9	88.9	1: 423
(Orlandi *et al.*, 1997)	36	1,956	1.81	5	6	83.3	1: 338
(Wapner *et al.*, 2003)	163	7,981	2.00	10	11	90.9	1: 755
All	587	87,062		96	103		1: 690

[1] includes both trisomy 18 and trisomy 13

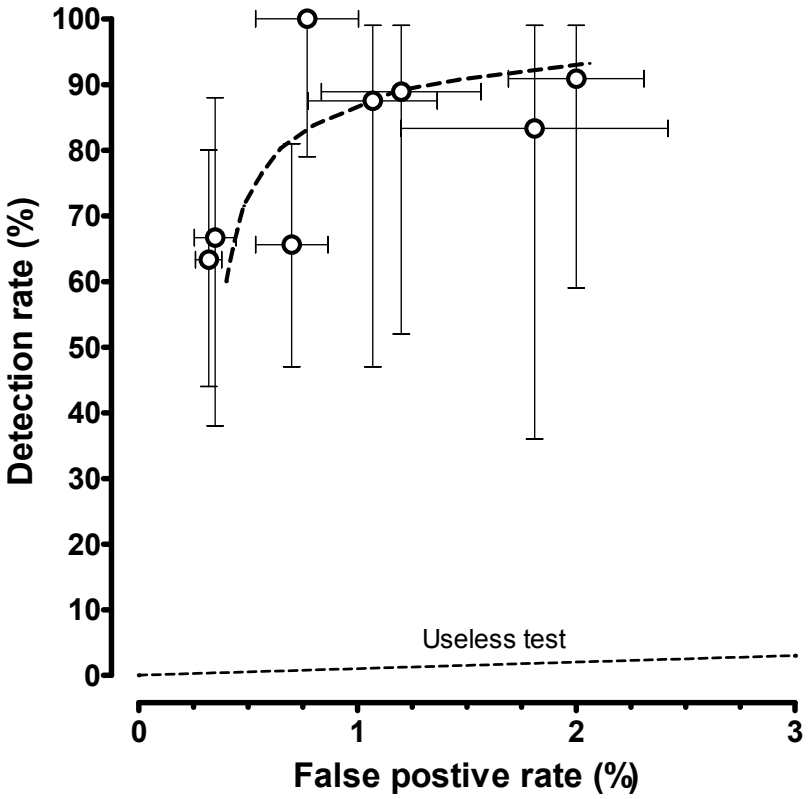

Figure 6.3-2. Trisomy 18 detection rate and associated false positive rate from eight demonstration studies of first trimester combined test. The open circle indicates the point estimate, with the thin lines indicating the corresponding 95% confidence interval. The dashed curve indicates the average performance over the eight studies. The dashed line indicated a 'useless' test, where the detection rate is equal to the false positive rate.

Figure 6.3-2 is a receiver operator characteristic (ROC) curve that summarizes the performance of the eight studies shown in Table 6.3-1. Since detection rate estimates are based on between 4 and 32 cases, they are broad compared to estimates of the false positive rates. This performance is not as high as that predicted by modeling (Table 6.1-1), where detection rates from 80% to 90% could be achieved with false positive rates below 0.5%. The most important observation, however, is the high false positive rates reported in the studies. Only two studies were below a 0.5% false positive rate and these were both published recently (Breathnach *et al.*, 2007; Jaques *et al.*,

2007). The most recent report was only slightly higher at 0.7% (Schaelike *et al.*, 2009). This might indicate that the earlier studies did not choose an optimal risk cut-off level. It is also clear that not all of pregnancies associated with a 'false positive' result have normal fetuses. Elevated NT measurements are clearly associated with heart defects and other abnormalities and this may be another reason why the modeled false positive rates should be viewed with caution. It is expected that the false positive rates in practice will be higher than the modeled rates by at least a few tenths of a percent. This was clearly demonstrated in the evaluation of second trimester serum-based testing for trisomy 18 (Section 3.11).

Excluded studies: To help clarify why studies were excluded, two examples were chosen for further examination. Excluding a study does not imply that it is a poor study. It only suggests that that study did not provide reliable data suitable to answer the question posed. The recent study from Denmark (Torring, 2009) provides an overview of 44,537 singleton pregnancies screened between 8 and 13 completed weeks. Over 80% of the data were collected at 8-10 weeks, outside of the window chosen for study. In additon, the author grouped trisomy 18, trisomy 13 and triploidy together. Lastly, there was no clear algorithm specified. Rather, this group of defects were identified by 'clincal data and biochemistry, or clinical data and nuchal translucency measurements'. The study with the largest number of cases (61) was from the UK (Nicolaides *et al.*, 2005) and it applied only a Down syndrome risk protocol and then reported detection rates only for the combination of trisomy 18 and 13.

Summary: Based on included studies, a combination of maternal age, PAPP-A, free β (or intact hCG) and NT measurements can detect 80% to 90% of trisomy 18 pregnancies in the late first trimester with fewer than 0.5% of the population having a screen positive result. Many of these women will also be screen positive for Down syndrome, as the marker patterns for the two abnormalities are similar; only the free β measurements differ (elevated for Down syndrome, but reduced for trisomy 18). Given the similarity of the pattern for trisomy 18 and trisomy 13, it is reasonable to assume that the majority of trisomy 13 pregnancies will also be identified as part of either the Down syndrome or trisomy 18 risk interpretations. Due to a lack of relatively unbiased data for NT measurements in trisomy 13 pregnancies, it is not possible to be more specific.