Chapter 3. Second trimester maternal serum markers for trisomy 18

3.1 Introduction and background

Introduction The aim of this chapter is to review the literature to answer a series of questions relating to second trimester serum-based testing for trisomy 18.

- Based on observational studies, what are the summary population parameters for maternal serum markers in second trimester trisomy 18 pregnancies, and do they agree with those in wide use today (Palomaki *et al.*, 1995)?
- Do trisomy 18 demonstration studies using AFP, uE3 and hCG confirm the screening performance of the risk-based model (Palomaki *et al.*, 1995), as defined by the false positive rate and positive predictive value?
- Are measurements of inhibin-A useful to add to the risk-based model? If so, describe the algorithm and model the expected increase in performance.
- Are there other serum markers that may be of use in a risk-based model? If so, describe the algorithm and model the expected increase in performance.

After a brief introduction to risk-based screening for Down syndrome (the model system upon which trisomy 18 testing is based), at least six separate markers will be examined in some detail (AFP, uE3, hCG, free β subunit of hCG, inhibin-A and PAPP-A). Additional markers will be included if relevant data are found to show that inclusion is warranted. A final parameter set will be created, and combinations of useful markers for predicting patient-specific risks will be explored. Results from demonstration studies of already implemented testing protocols will also be examined.

Methods When examining each of the six markers, a structured literature search (through 2009) was conducted and results restricted to studies of singleton unaffected pregnancies, and pregnancies affected by trisomy 18. Specific inclusion criteria for each analyte will be provided. A figure representing the publication date and number of trisomy 18 pregnancies will be presented to examine sizes of studies, types of studies and trends over time. These figures also make explicit the number and size of each study that did not meet the exclusion criteria. A summary estimate of the central value will be derived using a weighted random effects model on the median (or geometric mean) MoM value in the trisomy 18 pregnancies, after a logarithmic transformation. Studies may have provided only the median MoM, only the geometric mean (in the form of the mean MoM after a logarithmic transformation), or both. Tables will provide both

estimates. If one of them was not reported, it will be estimated using the other (*e.g.*, a missing median value will be assigned the antilog of the logarithmic mean). If feasible, analysis will be restricted to studies with at least a minimum number of affected pregnancies. This requirement helps protect the summary from publication biases, and acknowledges that small datasets often do not report the summary information needed for the proposed analyses (*e.g.,* logarithmic standard deviation). Often, the data for trisomy 18 is contained within a study focused on Down syndrome and because of this, it is possible that some reports that included relatively small number may be missed.

The summary effect size estimates (usually the median marker measurement in trisomy 18 pregnancies) from the larger studies will be examined for heterogeneity between studies using the Q-statistic (weighted sum of squared differences between the individual and overall effect size), the I^2 value (representing the percentage of variability not explained by random chance) and a corresponding two-sided p-value. Q tends to have low power with analyses having relatively few entries, while being too likely to show significant heterogeneity when analyzing a large number of studies. Interpretation of I^2 does not depend on the number of studies. In general, $I^2 < 25\%$ indicate limited heterogeneity, 26% to 49% moderate, and >50% high heterogeneity. When significant and/or high heterogeneity is identified, stratified analyses will be undertaken, when possible, to identify potential sources for that heterogeneity. The analysis of AFP, the first marker associated with aneuploidy, is provided in detail, but less detail is provided for the remaining markers that utilize the same analytic methodology.

Background. Measurements of serum markers are expressed as multiples of the median level in unaffected pregnancies. This concept was introduced as part of the First UK Collaborative Study (Wald *et al.*, 1977), due to large differences in standardization between AFP assays in use at collaborating centers. Median levels at 16 weeks' gestation, for example, varied by more than a factor of two. In addition, the median AFP levels increased with advancing gestational age, presenting another barrier to the use of a fixed mass unit cut-off level (*e.g.*, >100 ng/mL). Figure 3.1-1 provides an introduction into the computation of the multiple of the median (MoM) for an individual woman, using the median level found in the general population as the reference. The relative distributions are plotted rather than the absolute distributions, in order to demonstrate how likelihood ratios can be computed. Likelihood ratios are the relative increase, or decrease, in risk resulting from the woman having a specific marker measurement. The median was chosen, rather than the average, because of the effect an occasional high outlying value might have on the estimate of the central measure. Reporting results in

MoM is considered the 'common currency' that allows laboratories to compare values, regardless of their testing platform. Factors other than gestational age (*e.g.,* maternal weight, maternal race) can be accounted for by adjusting the MoM levels. Test results reported in MoM are also the basic component used in creating patient-specific risks for Down syndrome and other fetal anomalies.

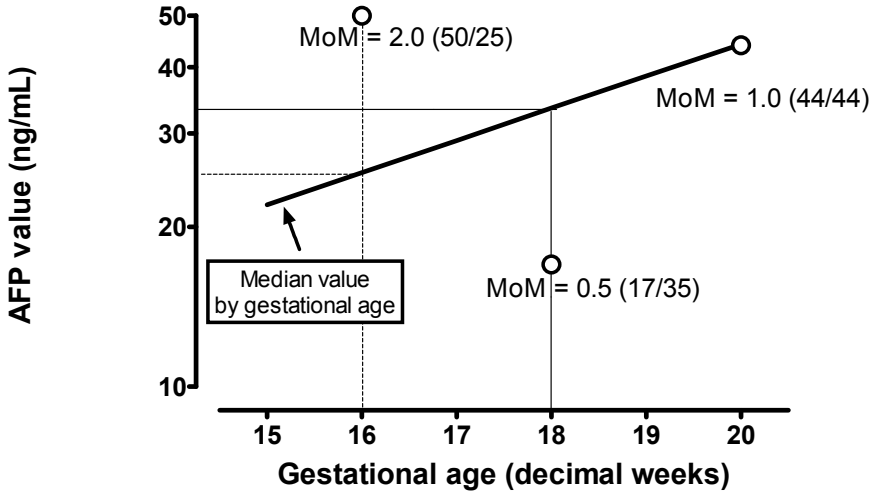

Figure 3.1-1. Sample calculations of an AFP multiple of the median (MoM). The solid thick line shows the AFP median levels increasing by gestational age by about 15% per week. The open circles represent the AFP measurements in three different women. The vertical lines indicate the gestational age, and the horizontal lines indicate the corresponding median AFP value. The MoM level is computed by taking the woman's value, divided by the corresponding median.

In the 1970s, reports began emerging about the relationship between open neural tube defects (ONTD) and levels of alpha-fetoprotein (AFP) in the amniotic fluid, and maternal serum AFP (Brock *et al.*, 1973; Field *et al.*, 1976; Brock and Sutcliffe, 1972). The methods for reporting analytic measurements in MoMs, computing detection and false positive rates, and computing patient specific risks were developed at this time (Wald *et al.*, 1977; Wald, 1976). Prenatal screening for ONTD became widespread in Europe and North America by the early 1980s (Ferguson-Smith *et al.*, 1978; Wald *et al.*, 1979; Burton *et al.*, 1983; Haddow *et al.*, 1983). In 1983, as part of routine prenatal care, a woman in

New York received a very low risk for ONTD due to an undetectable AFP result. She continued her pregnancy and delivered a trisomy 18 child. Later, she consulted with her physician to determine whether the low levels might be indicative of her baby's outcome. Her consistent probing prompted the original report by a New York group (Merkatz *et al.*, 1984) of reduced maternal serum AFP levels not only in trisomy 18 and 13 pregnancies, but also in 25 Down syndrome pregnancies.

Within months, the finding was confirmed for Down syndrome (Cuckle *et al.*, 1984), and existing prenatal screening programs for ONTD quickly introduced an additional interpretation of the maternal serum AFP test results for Down syndrome. Down syndrome was the initial target of screening rather than trisomy 18 because it is more common, and because long-term survival after birth places ongoing burdens on the family. Unlike ONTD screening, which relies on a fixed AFP MoM cut-off level (*e.g.*, $\geq$2.5 MoM), Down syndrome screening utilizes a patient-specific risk. This risk can be computed using the woman's age-specific risk for Down syndrome, multiplied by the AFP likelihood ratio (relative increase or decrease in risk). Figure 3.1-2 shows the overlapping distributions of AFP measurements in unaffected and Down syndrome pregnancies. The three vertical lines indicate the likelihood ratio corresponding to three different AFP test results. This methodology was published in 1987 and is still widely used (Palomaki and Haddow, 1987; Cuckle *et al.*, 1987). This same approach can be applied to testing for trisomy 18.

With Figure 3.1-2 as a guide, one can compute patient specific risks using the maternal age-related *a-priori*, or background risk as a starting point. In Down syndrome, for example, the risk at term for 20, 30 and 40 year old women are about 1:1500, 1:690 and 1:53, respectively (Morris *et al.*, 2002). If each of these women were to have an AFP value of 0.5 MoM, the corresponding likelihood ratio (LR) would be 4.1 for all three. This indicates a 4.1-fold increase to their age-related background risk, so the right hand side of the odds are divided by this factor to produce the post-test odds of about 1:370, 1:170 and 1:13, respectively. The two Gaussian curves cross at 0.79 MoM, indicating an LR of 1.0. This means that their post test odds are the same as their background odds (*i.e.,* 1:1500, 1:690 and 1:53). As a third example, consider the three women's risk at 1.5 MoM. The corresponding LR is 0.38, indicating reduced risk of having a fetus affected with Down syndrome. The three post-test odds would now be approximately 1:3900, 1:1800 and 1:40, respectively.

As an aside, the risk at term is used to indicate the possibility of delivering a viable baby with Down syndrome. When this risk is presented as a numeric value, it can be in the form of an odds, or as a probability. A 10% probability is equivalent to an odds of 1:9. That is, one Down syndrome birth to nine unaffected births, or one out of the 10 births, or 10%. The "1 in 10" form may be confusing, as some could read it as an odds, while others as a probability. Therefore, it would be best to avoid this format.

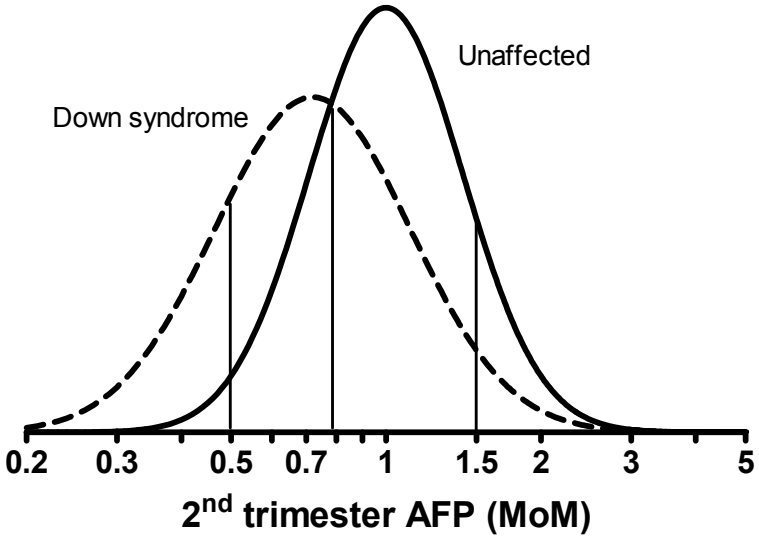

Figure 3.1-2. Overlapping maternal serum AFP measurements in unaffected and Down syndrome pregnancies. The solid curve shows the logarithmic Gaussian distribution of second trimester maternal serum AFP measurements in unaffected pregnancies (centered at 1.0 MoM). The dashed curve shows that the distribution in Down syndrome pregnancies is slightly lower (centered at 0.72 MoM) and somewhat broader (higher logarithmic standard deviation) than in unaffected pregnancies. The thin vertical lines at three AFP MoM levels provide examples of likelihood ratios of 4.0, 1.0 and 0.38, respectively.

One question to demonstrate understanding of the likelihood ratio is: "what happens to a woman's risk if the results were 'average' or 1.00 MoM? A common misconception is that the risk does not change. However, it can be easily seen from Figure 3.1-2 that the risk is actually decreased at 1.00 MoM, with a likelihood ratio of about 0.75. The more

useful a marker is, the higher the reduction in risk given the 'average' test result. The range of likelihood ratios provides some estimate of the power of the individual markers. Maternal age can also be treated as a 'screening test' in the same way as AFP measurements were in the previous paragraphs. The risk at term for a general population of women is about 1:600. If overlapping distributions of maternal age in women with, and without, a Down syndrome fetus were plotted, likelihood ratios could be generated at the three maternal ages provided earlier (20, 30 and 40 years of age). Although the curves would not be Gaussian, the ratio of the heights of the relative curves would still allow an approximate computation of the likelihood ratio. The expectation would be LRs of about 0.4, 0.9 and 11, respectively. When combinations of markers are used, LR in the hundreds or even thousands (or the reciprocals of these numbers) can be generated.

When combining multiple markers, each of which fits a logarithmic Gaussian distribution reasonably well, a multivariate Gaussian model can be used. These models are widely applied, and there is general agreement that they fit the data well (Wald *et al.*, 2003). In addition to the logarithmic mean and standard deviation for each marker in both affected and unaffected pregnancies, one also needs to have pair-wise correlation coefficients in both populations. Truncation limits need to be specified as well. Truncation limits are relatively extreme values for each of the markers that are used to help ensure the robustness of likelihood ratios generated by the model. A reasonable value for the lower truncation limit is the mean minus 2.5 standard deviations of the higher distribution (*i.e.*, the unaffected distribution for AFP measurements), while the upper limit would be the mean plus 2.5 standard deviations of the lower distribution (*i.e.*, Down syndrome pregnancies in this example). Several other factors need to then be considered before choosing a final set of truncation limits, including an inspection of the probability plots, lower limit of detection for the assay, similarity of the two standard deviations and overall range of possible likelihood ratios. Together, these means, standard deviations, correlation coefficients and truncation limits are called a set of population parameters. A 'set' of population parameters in a Gaussian model is used to calculate patient specific risks for Down syndrome, as well as several other outcomes (*e.g.*, trisomy 18, neural tube defects) using from one to five, or more, markers.

Table 3.1-1 provides a summary listing of the 14 studies supplying data for the analysis of AFP, uE3, hCG, the free β subunit of hCG, inhibin-A and PAPP-A in trisomy 18 pregnancies. As can be seen from the table, the majority of these studies report on more than one marker. The next sections examines each marker more closely.

Table 3.1-1. Studies of second trimester maternal serum markers and trisomy 18 included in the analyses

Included Publication	N	AFP	uE3	hCG	Free β	Inhibin-A	PAPP-A
(Lindenbaum *et al.*, 1987)	38	X					
(Zeitune *et al.*, 1991)	19	X					
(Spencer *et al.*, 1993)	52	X			X		
(Palomaki *et al.*, 1995)	89	X	X	X			
(Aitken *et al.*, 1996)	32	X		X	X	X	
(Leporrier *et al.*, 1996)	33		X	X			
(Wenstrom *et al.*, 1998)	13					X	
(Lambert-Messerlian *et al.*, 1998)	21	X	X	X		X	
(Sancken *et al.*, 1999)	30	X	X	X			
(Spencer *et al.*, 1999)	65	X			X		X
(Bersinger *et al.*, 1999)	20		X	X			X
(Kennedy *et al.*, 2000)	46	X		X			
(Muller *et al.*, 2002)	45	X			X		X
Included studies		10	5	7	4	3	3

3.2 Second trimester AFP measurements

As part of prenatal screening and/or diagnostic programs, an occasional trisomy 18 fetus is identified. By the late 1970s, it was well known that open ventral wall or open neural tube defects occurred in about 10 to 15% of trisomy 18 fetuses, and a proportion of these pregnancies were already being identified by ONTD screening programs. In the initial report of reduced AFP measurement and Down syndrome, the case that prompted further investigation was actually a trisomy 18 fetus (Merkatz *et al.*, 1984) not a Down syndrome fetus. By 1987, Lindenbaum and his colleagues in England and Finland reported that maternal serum AFP levels in 58 trisomy 18 pregnancies without open defects were reduced even more than that found for Down syndrome. Most had been collected as part of ONTD screening (Lindenbaum *et al.*, 1987). After removing these results, (the 20 trisomy 18 pregnancies with an associated open defect), they reported that a high proportion of the remaining affected fetuses could be detected as part of existing Down syndrome screening, without additional false positives. For example, using a fixed AFP MoM cut-off level of $\leq$0.5 MoM, 26% of Down syndrome and 34% of trisomy 18 fetuses (without ONTD) could be identified, along with about 8.4% of the unaffected pregnancies (Lindenbaum *et al.*, 1987). Since screening algorithms for Down syndrome screening at that time were based on only maternal age and serum AFP measurements, there was no reason to have a separate algorithm for trisomy 18, as these pregnancies would already be screen positive for Down syndrome.

The English literature through 2009 was searched for primary references regarding second trimester maternal serum AFP measurements in cytogenetically confirmed trisomy 18 pregnancies that were not part of a demonstration study for trisomy 18. A demonstration study is defined as a follow-up study of a testing protocol that includes a medical intervention to determine the effectiveness of the testing protocol in practice. This includes examining implementation issues such as uptake rates, and decision-making, and usually includes some sort of outcome follow-up. For example, a manuscript describing the reporting of trisomy 18 risks as part of a Down syndrome screening program (with pregnancy follow-up for those with a positive test result) would be considered a demonstration study. A Medline search (key words: trisomy 18, AFP, second trimester) yielded 36 references, 30 of which were not relevant. Based on the six relevant publications and a search of their reference lists, a total of 37 candidate publications were identified. Inclusion criteria were: 1) the peer-reviewed study was observational and would not be classified as a demonstration project, 2) the median or

geometric mean for trisomy 18 pregnancies was reported or could be computed from a published figure or table, 3) the gestational age range was reported and all, or nearly all, were in the range of 15 to 20 completed weeks' gestation, 4) information was available regarding how the pregnancy was identified for prenatal diagnosis (*e.g.,* live birth with an anomaly, serum screen positive, abnormal ultrasound) and 5) a reasonable number of observations were available. This latter inclusion criterion was added because of a large number of publications that focused on Down syndrome provided only sparse data for trisomy 18. A reasonable minimum number is 15 to 20 cases. All of the studies that focused on trisomy 18 are higher than this number, while most of the smaller studies that reported fragmented results are well below. Also extracted from the larger studies, when possible, were the logarithmic standard deviations and pair-wise correlation coefficients with other second trimester serum markers in both affected and unaffected pregnancies.

Three abstracts were identified and removed from further consideration (Subramanian, 1988; Darnule, 1990; Arab, 1988) as these were not considered peer-reviewed publications. One publication was in German (Dix *et al.*, 1988) and was also removed. None of these four publications contained more than five trisomy 18 pregnancies. An additional 22 publications had 14 or fewer observations and were not formally summarized (Staples *et al.*, 1991; Greenberg *et al.*, 1992; Canick *et al.*, 1990; Macri *et al.*, 1986; DiMaio *et al.*, 1987; Ashwood *et al.*, 1987; Doran *et al.*, 1986; Wenstrom *et al.*, 1998; Norgaard-Pedersen *et al.*, 1990; Palomaki *et al.*, 1992; Nebiolo *et al.*, 1990; Hershey *et al.*, 1985; Suchy and Yeager, 1990; Heyl *et al.*, 1990; Bogart *et al.*, 1987; Merkatz *et al.*, 1984; Huderer-Duric *et al.*, 2000; Crossley *et al.*, 1991; Cowchock, 1984). Several of these 22 publications might also have been excluded based on the majority of cases being identified as part of a demonstration study for trisomy 18. These are, therefore, likely to provide a biased estimate of the true levels of any serum marker in affected pregnancies. The 10 included studies were published between 1987 and 2002, and contained between 19 and 89 samples (Palomaki *et al.*, 1995; Muller *et al.*, 2002; Lindenbaum *et al.*, 1987; Kennedy *et al.*, 2000; Sancken *et al.*, 1999; Zeitune *et al.*, 1991; Spencer *et al.*, 1999; Aitken *et al.*, 1996; Lambert-Messerlian *et al.*, 1998; Spencer *et al.*, 1993). Of the 558 trisomy 18 pregnancies with maternal serum AFP measurements reported in the peer-reviewed literature, the 10 larger publications contained 436 observations (78% of the total) while the 22 smaller publications contained the remaining 122 observations (22% of the total). Figure 3.2-1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those studies included in the formal summaries (above dashed line) as well as those not included (below dashed line).

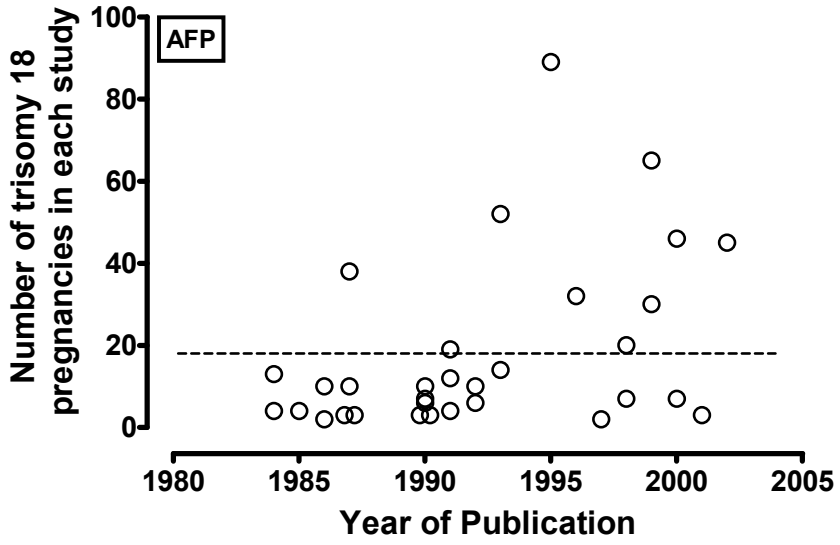

Figure 3.2-1. Maternal serum AFP measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Overall, 22 publications would be considered small (14 or fewer pregnancies) and the focus is usually not trisomy 18. A total of 10 studies included 19 or more trisomy 18 pregnancies (between 19 and 89) and would be considered large. In these larger studies, the focus was more often directed towards trisomy 18 rather than Down syndrome. These 10 studies form the basis of further analyses summarizing the second trimester serum AFP measurements in trisomy 18 pregnancies.

The AFP and trisomy 18 data are summarized in Table 3.2-1, ordered by the effect size (largest to smallest) as measured by the logarithmic mean. Stratified analyses can be performed using the covariates listed in the last three columns of Table 3.2-1. For example, do studies with earlier publication dates systematically differ from the later studies? This might occur because of improvements in assays or the routine inclusion of maternal weight adjustments. Also, do studies that included some proportion of samples identified via low maternal serum AFP measurements have a lower central estimate than those in which serum screening played no part in the identification of pregnancy outcomes?

Table 3.2-1. Maternal serum AFP measurements in trisomy 18 and unaffected pregnancies

		Trisomy 18 pregnancies			Unaffected pregnancies			Categories		
Reference	**GA**	**N**	**AFP Central Estimate (MoM)[a]**	**AFP Log SD**	**N**	**AFP MoM (Median)**	**AFP Log SD**	**>20% Screened by AFP**	**Pub After 1995**	**Log SD < 0.2113**
(Aitken *et al.*, 1996)	8-18	32[b]	-0.2757 (0.53)	NR	438	1.00	NR	Yes	Yes	NR
(Lambert-Messerlian *et al.*, 1998)	15-20	21	-0.2757 (0.53)	0.2117	105	1.00	0.1990	Yes	Yes	High
(Sancken *et al.*, 1999)	14-20	30	-0.2579 (0.73)	0.2215	29.081	1.00	0.1565	Yes	Yes	High
(Lindenbaum *et al.*, 1987)	early 2nd	38	-0.2170 (0.60)	0.3120	NR	NR	NR	No	No	High
(Muller *et al.*, 2002)	early 2nd	45	-0.2147 (0.61)	0.1830	15,000	1.00	NR	No	Yes	Low
(Spencer *et al.*, 1999)	14-19	65	-0.2050 (0.66)	0.2100	450	1.00	0.1750	Yes	Yes	Low
(Palomaki *et al.*, 1995)	13-22	89	-0.1970 (0.65)	0.2239	NR	NR	NR	Yes	No	High
(Zeitune *et al.*, 1991)	16-19	19	-0.1675 (0.68)	0.2201	133,045	1.00	0.1740	Yes	No	High
(Spencer *et al.*, 1993)	14-21	52	-0.1668 (0.71)	0.2037	6,661	0.99	0.1889	No	No	High
(Kennedy *et al.*, 2000)	14-21	46	-0.1565 (0.70)	0.1651	48,150	1.02	0.1415	No	Yes	Low
Summary		**436**	**-0.1830[d] (0.66[c])**	**0.1817[d]**		**1.00**	**0.1664**			
			95% CI (0.56-0.69)							

NR=not reported, GA=gestational age, log SD=logarithmic standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

[b] Only four of the 32 samples were collected under 15 weeks' gestation.

[c] After restriction to studies that did not include a high proportion of cases that had previous AFP testing (Figure 3.2-3).

[e] After regression analysis accounted for a significant temporal trend of reduced variances in affected pregnancies.

All studies in Table 3.2-1 confirmed the finding of trisomy 18 via karyotype, supplied the number of observations, provided results in multiples of the median, and reported central estimates (*e.g.*, median AFP MoM) after excluding trisomy 18 pregnancies that were also affected by an open defect (*e.g.*, open neural tube or ventral wall). None of the studies included twin pregnancies.

The logarithmic standard deviations for AFP measurements in trisomy 18 pregnancies uncomplicated by an open defect. The logarithmic standard deviation was reported for eight of the studies (Table 3.2-1). Methods for estimating the standard deviation varied, but using the interval from the 10^{th} to 90^{th} centiles divided by 1.282*2 was common, as a way to account for outlying values. In one study (Aitken *et al.*, 1996), the focus was on Down syndrome and inhibin-A measurements, and no estimate for the standard deviation in trisomy 18 pregnancies was possible. In another (Lambert-Messerlian *et al.*, 1998), the raw data were obtained from the authors, and the general formula for standard deviation was used (the study included 20 observations, and no outliers were present).

The earliest paper (Lindenbaum *et al.*, 1987) had the largest standard deviation, while a much later paper (Kennedy *et al.*, 2000) had the smallest. Figure 3.2-2 shows a plot of all nine reported estimates of the AFP variance in trisomy 18 pregnancies by year of publication. There is a clear and statistically significant trend towards reduced variance in the more recent studies. Although it was not possible to determine the exact cause, others have reported an important reduction in variance over time in AFP measurements in unaffected pregnancies (Wald *et al.*, 2000), and have applied that reduction in variance to existing Down syndrome parameters as a way to account for these changes. The most likely reason for this reduction is an improvement in AFP assays, but other possibilities include improvements in assigning gestational age and inclusion of routine adjustments for maternal weight.

A linear regression analysis (weighted by the square root of the number of observations) was applied to the log of the variance versus the year of publication using all 10 observations. The result was highly significant ($p<0.01$), but the slope was highly influenced by one quite high early observation (Lindenbaum *et al.*, 1987). The analysis was then rerun with this observation removed. The resulting regression line was not significant ($p=0.07$), but fitted the data well ($r=0.674$), is plausible, and confirms an earlier finding (Wald *et al.*, 2000). Given that no data were available after 2002, the regressed

variance at that time (0.3301) was taken to be a reasonable estimate for the standard deviation of current AFP measurements (expressed in MoM) for trisomy 18 pregnancies in the second trimester. This translates into a logarithmic standard deviation of 0.1817. This value is included in the summary line in Table 3.2-1.

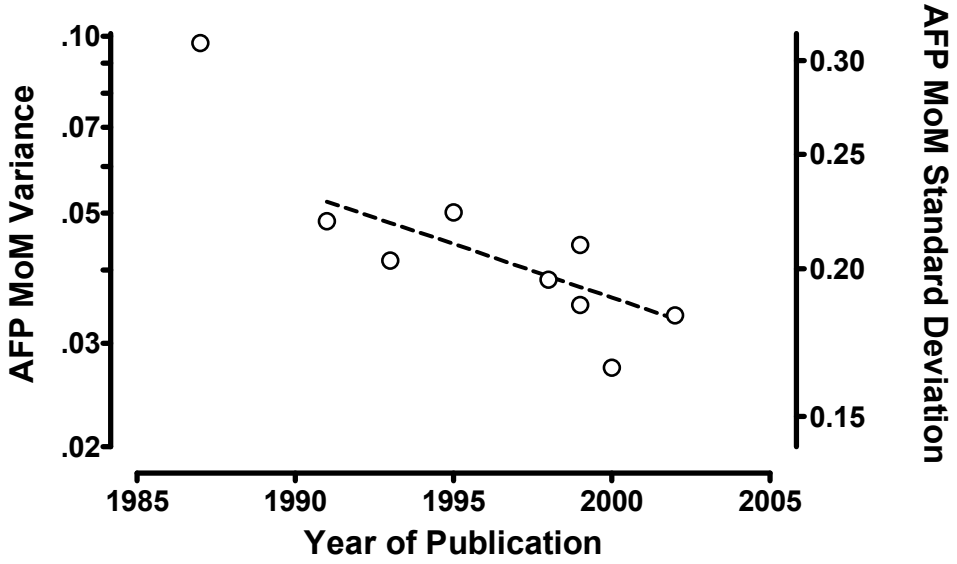

Figure 3.2-2. The variance of maternal serum AFP measurements in trisomy 18 pregnancies by year of publication. Each of the nine open circles shows an estimate from one of the large studies included in the analysis. The dashed line shows the results of a weighted log-linear regression analysis on the eight observations between 1991 and 2002.

The central estimate of maternal serum AFP in trisomy 18 pregnancies. All studies reported the median (and/or logarithmic mean) AFP MoM value in the trisomy 18 pregnancies. As described previously, all but one reported a logarithmic standard deviation (or a value was computed from the reported raw data). From this value and the number of cases, the standard error was computed. For the study with a missing logarithmic standard deviation, the regressed estimate of 0.1817 was used (interpolated for Figure 3.2-2, Table 3.2-1). The data were analyzed using a random effects model (Comprehensive Meta-Analysis V2.2, Biostat**,** Englewood, NJ). In addition to an estimate of the central value (*i.e.,* median, geometric mean), the model also was used to examine sources of heterogeneity. Figure 3.2-3 shows data for

each of the 10 studies. The weighted central estimate for the logarithmic mean is 0.209 and is shown in the last line in the forest plot. This corresponds to a median value of 0.62 MoM. The test for heterogeneity was negative (Q=15, p=0.07, I^2=43%), but the high I^2 value is suggestive of underlying variability. Among the excluded articles, 122 observations were available, and the weighted median value was 0.65 MoM. Even these more limited data were less well characterized and the data less reliable than those derived from larger, more focused studies.

Analyzing sources of heterogeneity. The last three columns of Table 3.2-1 provide stratifications of the data by three categorical variables: screened samples (none or fewer than 20% of samples identified as part of a serum screening program versus 20% or more identified via screening), logarithmic standard deviation (SD), and year of publication (prior to 1995 versus 1995 and later). Using a mixed-effects model, only one of these potential confounders appeared to be helpful in explaining the possible heterogeneity. Figure 3.2-3 shows the stratification by method of identifying the sample for diagnosis (ID). If trisomy 18 cases, even a proportion of the cases, were identified because of low maternal serum AFP measurements (*i.e.*, screen positive for Down syndrome using maternal age and low AFP), one might expect that those studies would have a lower AFP median than the remaining studies that identified cases only through abnormal ultrasound, or after a live birth. The six studies including pregnancies identified via serum screening had a median value of 0.59 MoM (95% CI 0.55 to 0.64), with the remaining four studies at 0.66 MoM (0.61 to 0.71) (overall, the estimate is 0.62 MoM, 95% CI 0.56 to 0.69). Although suggestive, the test for heterogeneity between the two groups is still negative (Q=3.6, p=0.06). However, given the plausibility of this finding, the higher estimate of 0.66 (95% CI 0.61-0.71, logarithmic mean of -0.1830) will be used as the summary estimate for the central value for AFP measurements in trisomy 18 pregnancies.

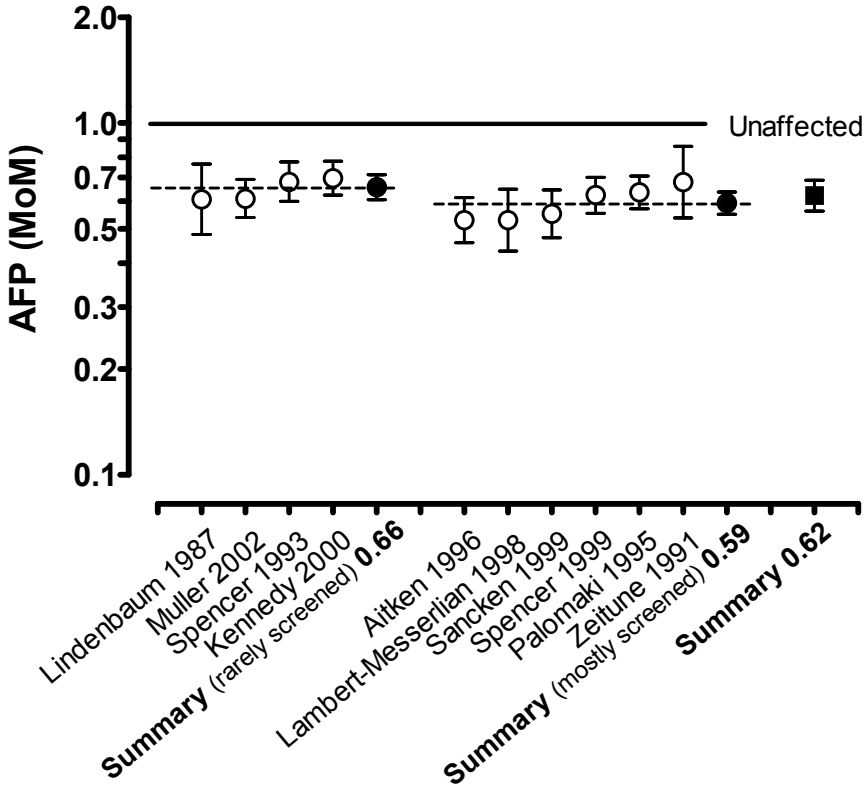

Figure 3.2-3. Maternal serum AFP measurements in trisomy 18 pregnancies. The central estimate (usually logarithmic mean) and associated 95% confidence intervals are shown for the 10 studies reporting AFP measurements from at least 19 pregnancies affected with trisomy 18 (Table 3.2-1). The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The four studies on the left represent those in which samples were rarely identified via serum testing (see Table 3.2-1). The next six studies were often (more than 20% of the time) taken from pregnancies identified via serum testing for Down syndrome. The overall summary estimate is 0.62, but the reduced effect size of 0.59 MoM (95% CI 0.55 to 0.64) will be used in modeling.

3.3 Second trimester uE3 measurements

There are fewer studies reporting uE3 measurements in trisomy 18 pregnancies. A Medline search (trisomy 18, uE3, unconjugated estriol, second trimester) identified 46 publications. Inclusion criteria were similar to those described earlier for AFP measurements. Three of these contained relevant information (Suzumori *et al.*, 1997; Palomaki *et al.*, 1995; Lambert-Messerlian *et al.*, 1998). Examining references in these papers identified additional potentially relevant publications. One abstract was excluded (Darnule *et al.*, 1990). Ten studies contained 15 or fewer observations and were not included (Barkai *et al.*, 1993; Greenberg *et al.*, 1992; Staples *et al.*, 1991; Canick *et al.*, 1990; Crossley *et al.*, 1993; Heyl *et al.*, 1990; Norgaard-Pedersen *et al.*, 1990; Kim *et al.*, 2001; Suzumori *et al.*, 1997; Palomaki *et al.*, 1992). The largest of these studies (Barkai *et al.*, 1993) identified 15 cases that were, in part, screen positive using a triple marker algorithm (AFP, uE3 and hCG) and, therefore, likely to not represent an unbiased estimate. The five remaining publications (Palomaki *et al.*, 1995; Sancken *et al.*, 1999; Leporrier *et al.*, 1996; Bersinger *et al.*, 1999; Lambert-Messerlian *et al.*, 1998) formed the basis of the following uE3 analyses. Overall, 280 observations were identified in all 15 peer-reviewed publications. The five publications included for analysis contained 201 of those observations (72%). The 10 publications with smaller numbers of observations contained an additional 79 observations (28%). Figure 3.3-1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those studies included (filled circles) and those not formally summarized (open circles).

Two studies (Palomaki *et al.*, 1995; Sancken *et al.*, 1999) noted that the uE3 measurements were not significantly different in pregnancies with open defects (spina bifida and ventral wall). For this reason, trisomy 18 pregnancies with open defects were included. The results of these analyses are presented below.

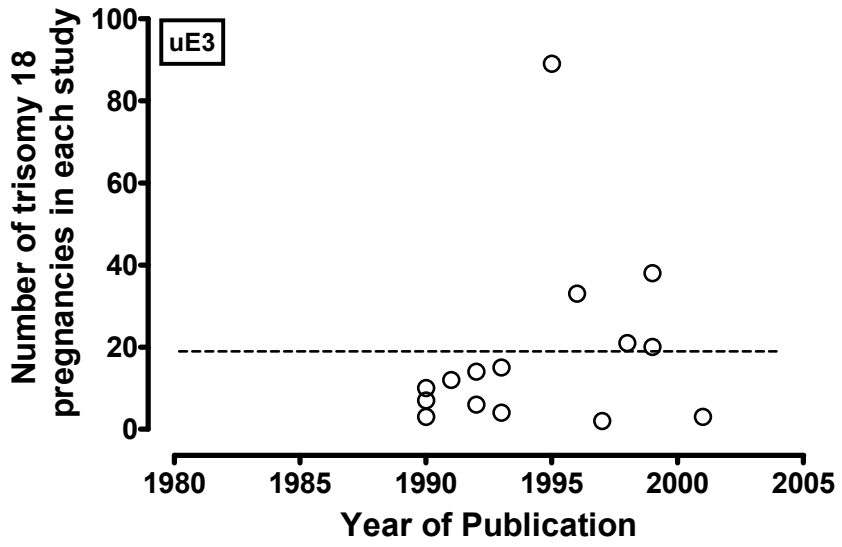

Figure 3.3-1. Publications reporting maternal serum uE3 measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Five studies were considered large (between 20 and 89 affected pregnancies) and these were used in the analysis of uE3 measurements in trisomy 18 pregnancies.

The central estimate of serum uE3 measurements in trisomy 18 pregnancies. All five studies reported the median and/or logarithmic mean uE3 MoM value for the trisomy 18 pregnancies studied, along with a logarithmic standard deviation. A random effects model was used to estimate the median uE3 MoM (logarithmic mean) in trisomy 18 pregnancies (Figure 3.3-2). The pooled logarithmic mean was found to be -0.4448, corresponding to a median value of 0.36 MoM. The test for heterogeneity was not statistically significant, but did indicate the possibility of between study differences ($Q=8.7$, $p=0.07$, $I^2=54\%$). Among the 79 usable observations from excluded studies, the weighted median MoM was 0.45. The discrepancy may be at least partially explained by the majority of excluded studies being performed between 1990 and 1995, when the uE3 assays were not optimized for the levels found in the second trimester.

Table 3.3-1. Maternal serum uE3 measurements in trisomy 18 and unaffected pregnancies

Reference	GA	Trisomy 18 pregnancies			Unaffected pregnancies		
		N	uE3 Central Estimate (MoM)[a]	uE3 Log SD	N	Median uE3 (MoM)	uE3 Log SD
(Sancken *et al.*, 1999)	14-20	38	-0.5429 (0.31)	0.4213	29,081	1.00	0.1565
(Bersinger *et al.*, 1999)	15-20	20	-0.5151 (0.32)	0.2677	40	1.10	0.1719
(Leporrier *et al.*, 1996)	14-23	33	-0.4706 (0.37)	0.2619	3,000	1.00[b]	0.1625
(Palomaki *et al.*, 1995)	13-22	89	-0.3991 (0.43)	0.2938	NR	NR	NR
(Lambert-Messerlian *et al.*, 1998)	15-20	21	-0.3372 (0.46)	0.2655	105	1.00	0.1742
Summary		**201**	**-0.4448 (0.36)**	**0.2817**			**0.1591**
			95% CI (0.31-0.42)				

NR = not reported, GA = gestational age, SD = standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

[b] Information about uE3 measurements in control pregnancies reported in a separate reference (Herrou *et al.*, 1992).

The logarithmic standard deviations for uE3 measurements in trisomy 18 pregnancies. In one study (Lambert-Messerlian *et al.*, 1998), the raw data were obtained from the author, and the general formula for standard deviation was used (only 20 observations were included in that study, and no outliers were identified). A previously published one-pass methodology (Palomaki *et al.*, 2007) was used to identify and trim outlying variances. Briefly, an F-value was computed using each individual variance divided by the pooled variance. If the F-value was significant, that value was trimmed and the summary recomputed. The weighted pooled standard deviation, prior to trimming, was 0.3131. The logarithmic standard deviation from one study (Sancken *et al.*, 1999) was identified as being an outlier, with a p-value of 0.01 (F=1.81, df=38). After removal, the trimmed pooled logarithmic standard deviation was 0.2817. These standard deviations were not subjected to a temporal analysis for three reasons. The timeframe over which the four studies took place was short (only four years), and the assay methodologies did not change over that time period. Lastly, the four estimates remaining after trimming are similar.

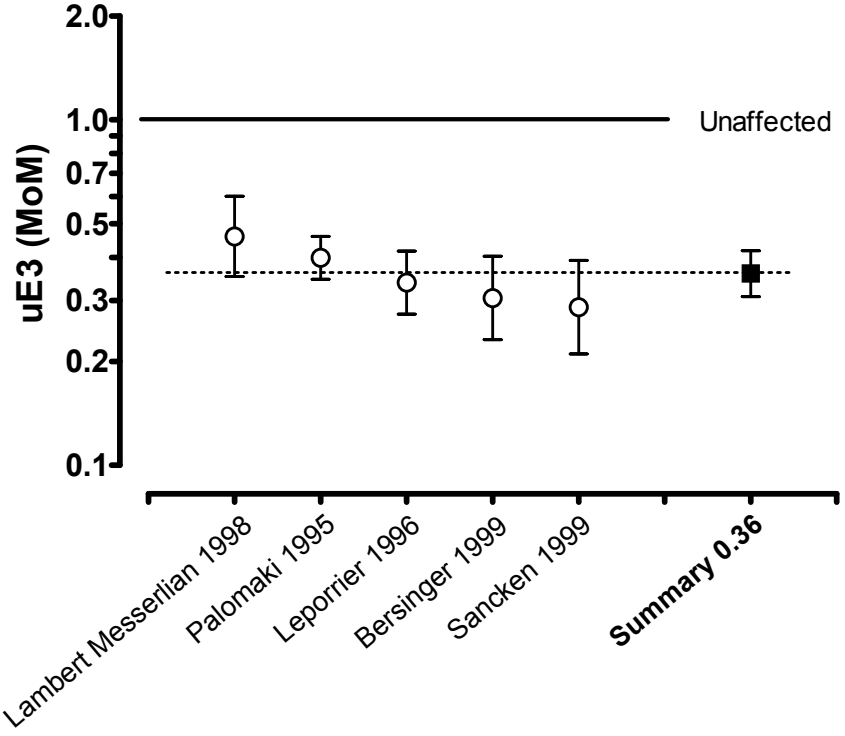

Figure 3.3-2. Maternal serum uE3 measurements in trisomy 18 pregnancies. The central estimate (usually logarithmic mean) and 95% confidence intervals are shown for the five studies reporting uE3 measurements from at least 19 pregnancies with trisomy 18. The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The summary logarithmic mean is -0.4448 (uE3 median of 0.36 MoM, 95% CI 0.31-0.42).

3.4 Second trimester hCG measurements

A Medline search (trisomy 18, hCG, human chorionic gonadotropin, second trimester) identified 92 publications. Six of these contained relevant information (Palomaki *et al.*, 1995; Kim *et al.*, 2001; Sancken *et al.*, 1999; Suzumori *et al.*, 1997; Canick *et al.*, 1990; Lambert-Messerlian *et al.*, 1998). Examining references in these papers identified additional relevant papers and abstracts. Inclusion criteria were similar to those described earlier for AFP measurements. Two abstracts were excluded (Blitzer, 1991; Arab, 1988). Twelve studies contained 14 or fewer observations and were also excluded (Greenberg *et al.*, 1992; Bartels *et al.*, 1990; Canick *et al.*, 1990; Norgaard-Pedersen *et al.*, 1990; Palomaki *et al.*, 1992; Nebiolo *et al.*, 1990; Kim *et al.*, 2001; Heyl *et al.*, 1990; Suchy and Yeager, 1990; Bogart *et al.*, 1987; Suzumori *et al.*, 1997; Crossley *et al.*, 1991). There were seven larger studies (20 or more observations) that met inclusion criteria (Palomaki *et al.*, 1995; Sancken *et al.*, 1999; Kennedy *et al.*, 2000; Leporrier *et al.*, 1996; Bersinger *et al.*, 1999; Aitken *et al.*, 1996; Lambert-Messerlian *et al.*, 1998), and these formed the basis of the hCG analyses. Overall, a total of 334 observations were identified in the 19 publications. The seven publications with 20 or more samples contained a total of 248 observations (74%), while the remaining 12 publications contained 86 observations (26%). Figure 3.4 -1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those studies included (filled circles) and those not formally summarized (open circles).

Two studies (Palomaki *et al.*, 1995; Sancken *et al.*, 1999) noted that the hCG measurements were not different in pregnancies with open defects. For this reason, trisomy 18 pregnancies with open defects were included. None of the studies included twin pregnancies. The results of these analyses are presented below.

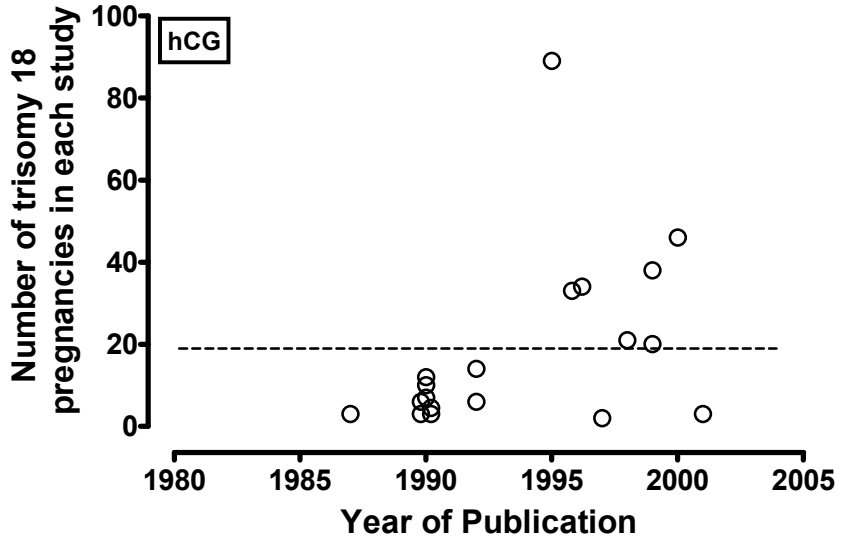

Figure 3.4-1. Publications reporting maternal serum hCG measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Seven studies were considered large (between 20 and 89 affected pregnancies) and these were used in the analysis of hCG measurements in trisomy 18 pregnancies.

The central estimate of serum hCG measurements in trisomy 18 pregnancies.

All seven studies reported the median and/or logarithmic mean uE3 MoM value, and all but one (Aitken *et al.*, 1996) also reported a logarithmic standard deviation (the trimmed pooled estimate of 0.3561 was used). A random effects model was used to estimate the median hCG MoM (logarithmic mean) in trisomy 18 pregnancies (Figure 3.4-2). The pooled logarithmic mean was -0.4601, corresponding to a median value of 0.35 MoM. The test for heterogeneity was significant (Q=18, p=0.007, I^2=66%). At least some of this heterogeneity is due to the estimate in the smallest study, consisting of 20 observations (Lambert-Messerlian *et al.*, 1998), where the median is 0.29. However, the logarithmic mean corresponds to a median of 0.22. Were the median of 0.29 to be used in the summary analysis instead, the test for heterogeneity would be reduced (Q=12, p=0.06, I^2=51%). Interestingly, this would have little effect on the summary MoM (from 0.35 to 0.36), and only a slight change in the logarithmic mean (from -0.4601 to -0.4442). For these reasons, the original summary of 0.35 MoM will be used. The weighted median MoM from the 86 usable samples was 0.29. This lower median MoM may be due to several of the larger excluded studies being demonstration studies, which would be expected to produce a lower (biased) estimate.

Table 3.4-1. Maternal serum hCG measurements in trisomy 18 and unaffected pregnancies

Reference	GA	Trisomy 18 pregnancies			Unaffected pregnancies		
		N	hCG Central Estimate (MoM)[a]	uE3 Log SD	N	Median uE3 (MoM)	uE3 Log SD
(Lambert-Messerlian *et al.*, 1998)	15-20	21	-0.6576 (0.29)	0.3550	100	1.00	0.2494
(Sancken *et al.*, 1999)	14-20	38	-0.5528 (0.28)	0.4734	29,081	1.05	0.2870
(Aitken *et al.*, 1996)	8-18	34	-0.5229 (0.30)	NR	438	1.03	0.2196
(Palomaki *et al.*, 1995)	13-22	89	-0.4396 (0.36)	0.3772	NR	NR	NR
(Bersinger *et al.*, 1999)	15-20	20	-0.4278 (0.41)	0.3870	40	0.99	0.2168
(Kennedy *et al.*, 2000)	14-21	46	-0.3659 (0.43)	0.3245	48,150	1.01	0.2250
(Leporrier *et al.*, 1996)	14-23	33	-0.3294 (0.50)	0.3177	3,000	1.00[a]	0.2619
Summary		**248**	**-0.4123 (0.39)**	**0.3561**			
			95% CI (0.29-0.41)				

NR = not reported, GA = gestational age, SD = standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

[b] Information about hCG measurements in control pregnancies from another publication (Herrou *et al.*, 1992).

The logarithmic standard deviation for hCG measurements The logarithmic standard deviation reported by the authors was used for six of the included studies. In another study (Aitken *et al.*, 1996), the focus was on Down syndrome and inhibin-A measurements, and no estimate for the standard deviation was possible. The weighted pooled standard deviation, prior to trimming, was 0.3768. The logarithmic standard deviation from one study (Sancken *et al.*, 1999) was identified as being an outlier, with a p-value of 0.04 (F=1.58). After this study was removed, the trimmed pooled logarithmic standard deviation was 0.3561.

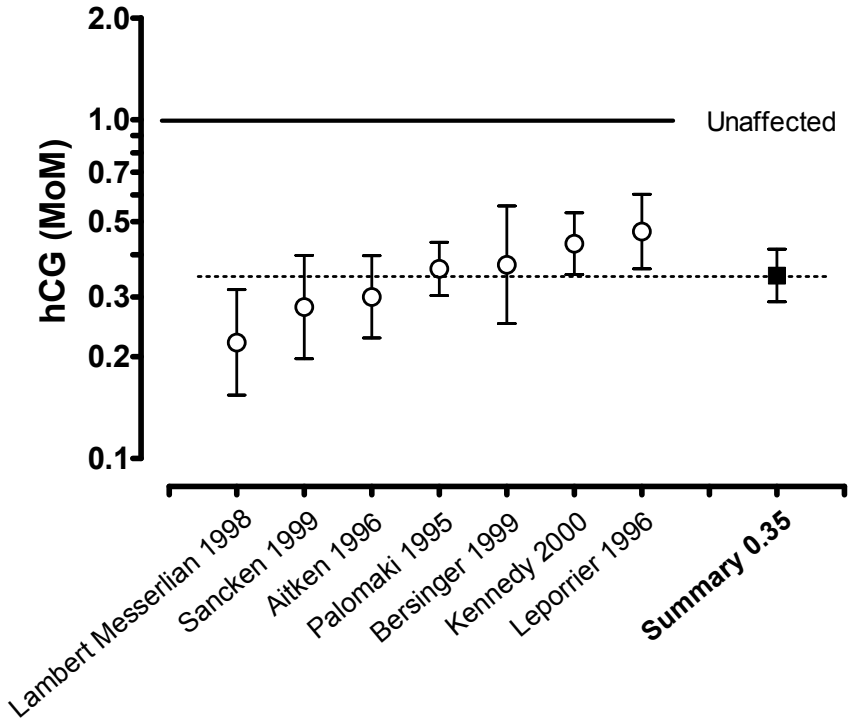

Figure 3.4-2. Maternal serum hCG measurements in trisomy 18 pregnancies. The central estimate (usually logarithmic mean) and associated 95% confidence intervals are shown for the seven studies reporting hCG measurements from at least 20 pregnancies affected with trisomy 18. The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The summary logarithmic mean is -0.4601 (median of 0.35 MoM, 95% CI 0.29-0.41).

3.5 Second trimester free β hCG

A Medline search (trisomy 18, free β hCG, human chorionic gonadotropin, second trimester) identified 33 publications. Inclusion criteria were similar to those described earlier for AFP measurements. Two of these contained relevant information (Muller *et al.*, 2002; Spencer *et al.*, 1999). Examining references in these papers identified additional relevant papers. Two studies included 12 or fewer observations (Staples *et al.*, 1991; Wenstrom *et al.*, 1998) and were excluded. There were four large studies (20 or more observations) that met inclusion criteria (Muller *et al.*, 2002; Spencer *et al.*, 1999; Spencer *et al.*, 1993; Aitken *et al.*, 1996), and these formed the basis of the free β hCG analyses. Overall, a total of 218 observations were identified in the six publications. The four publications included for analysis contained a total of 199 observations (91%), while the two publications with smaller numbers of observations contained the remaining 19 observations (9%). Figure 3.5-1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those studies included (filled circles) and those not formally summarized (open circles).

The logarithmic standard deviations for free β hCG measurements in trisomy 18 pregnancies. The logarithmic standard deviation reported by the authors was used for three of the included studies. In one study (Aitken *et al.*, 1996), the focus was on Down syndrome and inhibin-A measurements and no estimate for the standard deviation was possible. The weighted pooled standard deviation, prior to trimming was 0.4062. No outliers were identified.

The central estimate of serum free β hCG measurements in trisomy 18 pregnancies. All four included studies reported the median and/or logarithmic mean free β hCG MoM value for the trisomy 18 pregnancies studied, and all but one (Aitken *et al.*, 1996) also reported a logarithmic standard deviation. For this study, the trimmed pooled estimate of 0.4062 was used to estimate standard error. The standard error of the logarithmic mean was then computed, using the number of reported cases. A random effects model was used to estimate the median free β hCG MoM (logarithmic mean) in trisomy 18 pregnancies (Figure 3.5-2). The pooled logarithmic mean was -0.6203, corresponding to a median value of 0.24 MoM. The test for heterogeneity was significant ($Q=17$, $p=0.001$, $I^2=82\%$). Among the excluded studies, only one provided an estimate of the median at 0.31 MoM based on 12 observations (Staples *et al.*, 1991).

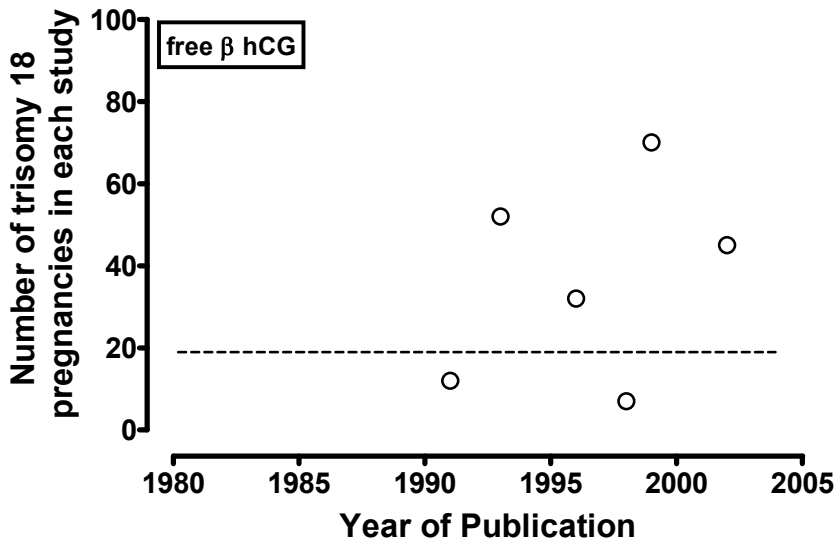

Figure 3.5-1. Publications reporting maternal serum free β hCG measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Four studies were considered large (between 32 and 70 affected pregnancies) and these were used in the analysis of free β measurements in trisomy 18 pregnancies.

Table 3.5-1. Maternal serum free β hCG measurements in trisomy 18 and unaffected pregnancies

Reference	GA	Trisomy 18 pregnancies			Unaffected pregnancies		
		N	Free β Central Estimate (MoM)[a]	Free β Log SD	N	Median free β (MoM)	Free β Log SD
(Aitken *et al.*, 1996)	8-18	32	-0.8539 (0.14)	NR	112	1.03	NR
(Muller *et al.*, 2002)	early 2nd	45	-0.6198 (0.24)	0.4010	15,000	1.00	NR
(Spencer *et al.*, 1999)	14-19	70	-0.5300 (0.33)	0.3810	450	1.00	0.2600
(Spencer *et al.*, 1993)	14-21	52	-0.5025 (0.37)	0.4346	6,661	0.99	0.2583
Summary		**199**	**-0.6203 (0.24)**	**0.4062**			
			95% CI (0.17-0,33)				

NR = not reported, GA = gestational age, SD = standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

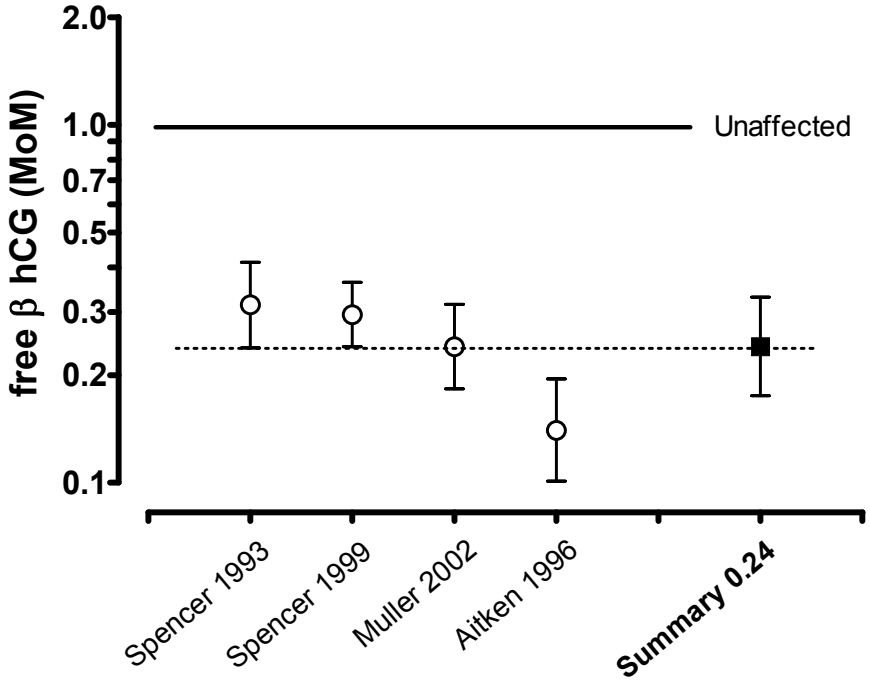

Figure 3.5-2. Maternal serum free β hCG measurements in trisomy 18 pregnancies. The central estimate (usually logarithmic mean) and associated standard errors are shown for the four studies reporting free β hCG measurements from at least 20 pregnancies affected with trisomy 18. The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The summary logarithmic mean is -0.6203 (free β hCG median of 0.24 MoM, 95% CI 0.17-0.33).

3.6 Second trimester inhibin-A measurements

A Medline search (trisomy 18, inhibin-A, second trimester) identified 15 publications. Inclusion criteria were similar to those described earlier for AFP measurements. Five of these contained relevant information (Hsu et al., 2003; Yoshida et al., 2000; Watanabe et al., 2002; Lambert-Messerlian et al., 1998; Aitken et al., 1996). Examining references in these papers identified additional relevant papers. Four studies included 13 or fewer observations (Wenstrom et al., 1998; Watanabe et al., 2002; Hsu et al., 2003; Yoshida et al., 2000). Because of the small number of studies, these will be included in the data listings but not in the subsequent analyses. There were only two larger studies (21 and 32 observations) that met inclusion criteria (Lambert-Messerlian et al., 1998; Aitken et al., 1996), and these formed the basis of the inhibin-A analyses. Overall, 78 observations were identified in the six publications. Figure 3.6-1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those two larger studies (filled circles) and the four smaller ones (open circles).

The logarithmic standard deviations for inhibin-A measurements in trisomy 18 pregnancies The logarithmic standard deviation reported by the authors was used for the two largest studies. The weighted pooled estimate is 0.2897.

The central estimate of serum inhibin-A measurements in trisomy 18 pregnancies The standard error of the logarithmic mean was computed for the two largest studies using the number of reported cases and the reported standard deviation. A random effects model was used to estimate the median inhibin-A MoM (logarithmic mean) in trisomy 18 pregnancies (Figure 3.6-2). The pooled logarithmic mean was found to be -0.0392, corresponding to a median value of 0.91 MoM. Neither study found a statistically significant difference, nor was the summary finding significantly different from 1.0 MoM (p=0.3). The four smaller studies (Wenstrom *et al.*, 1998; Watanabe *et al.*, 2002; Hsu *et al.*, 2003; Yoshida *et al.*, 2000) are consistent with no effect. The largest (Wenstrom *et al.*, 1998) summarized results from 13 affected pregnancies and reported 'no discrimination', but no associated parameters or summary statistics. The three smallest studies all found point estimates above 1.00 MoM.

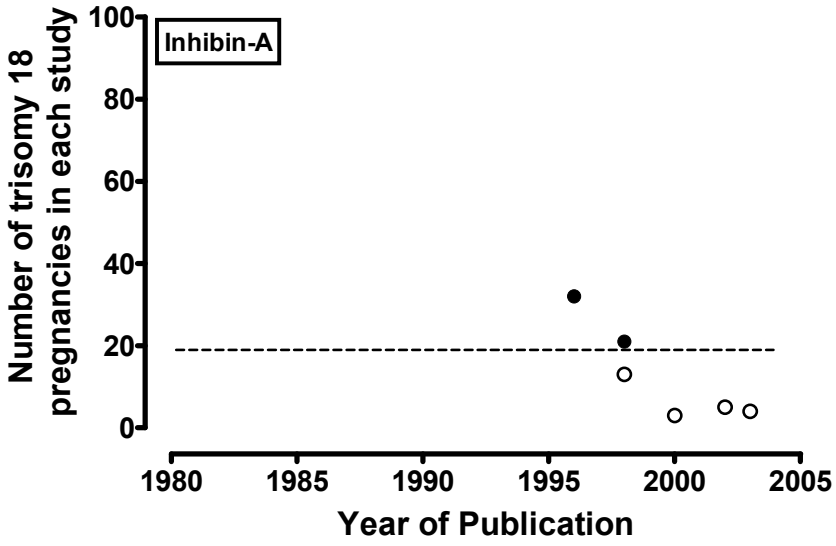

Figure 3.6-1. Publications reporting maternal serum inhibin-A measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Two studies were considered large (between 21 and 32 affected pregnancies) and these were used in the analysis of inhibin-A measurements in trisomy 18 pregnancies. Given the small number of studies and observations, all studies will be reviewed.

Table 3.6-1. Maternal serum inhibin-A measurements in trisomy 18 and unaffected pregnancies

Reference	GA	Trisomy 18 pregnancies			Unaffected pregnancies		
		N	Inhibin-A Central Estimate (MoM)[a]	Inhibin-A Log SD	N	Median Inhibin-A (MoM)	Inhibin-A Log SD
(Lambert-Messerlian *et al.*, 1998)	15-20	21	-0.0560 (0.88)	0.3406	100	1.00	0.2634
(Aitken *et al.*, 1996)	15-18	32	-0.0340 (0.99)	0.2455	112	1.00	0.2967
(Wenstrom *et al.*, 1998)	14-20	13	NR (NR)[b]	NR	450	NR	NR
(Yoshida *et al.*, 2000)	15-21	3	0.0253 (1.06)	NR	71	1.00	NR
(Watanabe *et al.*, 2002)	15-17	5	0.0710 (1.20)	0.0695	56	1.00	NR
(Hsu *et al.*, 2003)	14-22	4	-0.0990 (1.04)	0.6210	160	1.00	0.2343
Summary		**78**	**-0.0392 (0.91)**	**0.2897**			
			95% CI 0.77-1.09				

NR = not reported, GA = gestational age, SD = standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

[b] Reported there was 'no discrimination'.

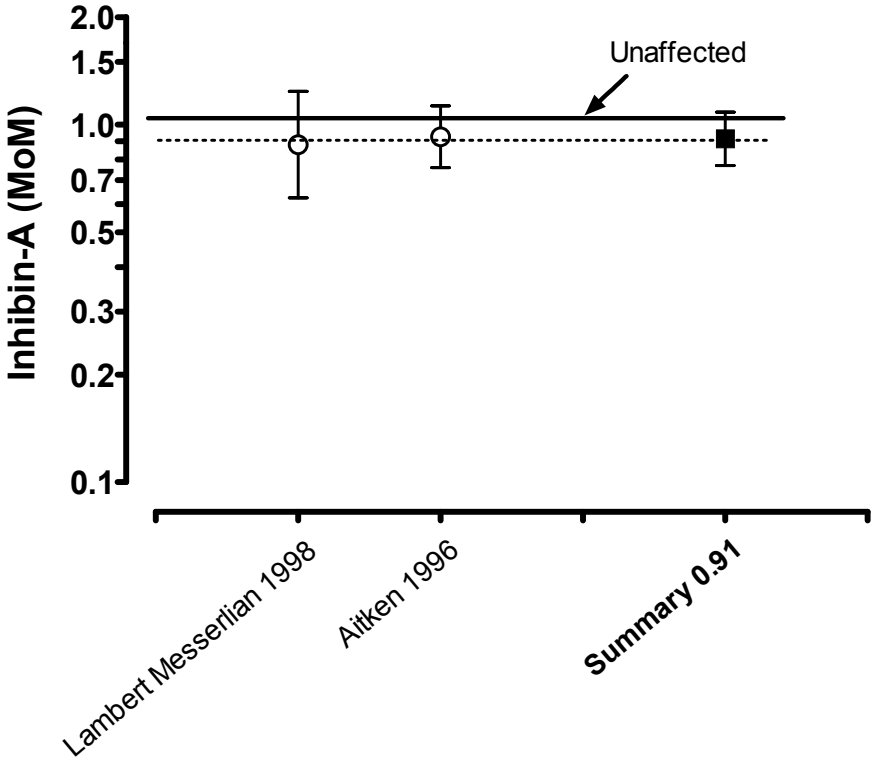

Figure 3.6-2. Maternal serum inhibin-A measurements in trisomy 18 pregnancies. The central estimate (usually logarithmic mean) and associated 95% confidence intervals are shown for the two studies reporting inhibin-A measurements from at least 20 pregnancies affected with trisomy 18. The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The summary logarithmic mean is -0.0392 (median of 0.91 MoM, 95% CI 0.77-1.09).

3.7 Second trimester PAPP-A measurements

Measurements of pregnancy associated plasma protein-A (PAPP-A) are known to be the most effective serum marker in the late first trimester for Down syndrome (and trisomy 18). However, PAPP-A measurements have been shown to be of little value in the second trimester for Down syndrome. For this reason, the expectation is for only a limited number of studies of second trimester PAPP-A and trisomy 18.

A Medline search (PAPP-A, serum, second trimester, trisomy 18) identified 26 publications. Inclusion criteria were similar to those described earlier for AFP measurements. Four of these contained relevant information (Muller *et al.*, 2002; Spencer *et al.*, 1999; Bersinger *et al.*, 1999; Watanabe *et al.*, 2002). Examining references in these papers identified no additional relevant publications. One of the four studies included only five observations and was excluded (Watanabe *et al.*, 2002). That left three large studies (20 or more observations) that met inclusion criteria, and these formed the basis of the PAPP-A analyses (Bersinger *et al.*, 1999; Muller *et al.*, 2002; Spencer *et al.*, 1999). Overall, 140 observations were identified in the four publications. The three publications included for analysis contained a total of 135 observations (96%), while the smaller publication contained the remaining 5 observations (4%). Figure 3.7-1 provides a graphical summary of year of publication versus numbers of trisomy 18 observations reported, and notes those studies included (filled circles) and those not formally summarized (open circles).

The logarithmic standard deviations for PAPP-A measurements in trisomy 18 pregnancies The logarithmic standard deviation was reported by all authors; the pooled estimate was 0.3894, with no outliers identified.

The central estimate of serum PAPP-A measurements in trisomy 18 pregnancies All three studies reported the median and/or logarithmic mean PAPP-A MoM value for the trisomy 18 pregnancies studied. A random effects model was used to estimate the median PAPP-A MoM (logarithmic mean) in trisomy 18 pregnancies (Figure 3.7-2). The pooled logarithmic mean was -0.9871, corresponding to a median value of 0.10 MoM. The test for heterogeneity was borderline significant ($Q=5.7$, $p=0.06$, $I^2=65\%$). Although there is some argument for heterogeneity, all three studies show an impressively low level of PAPP-A among trisomy 18 pregnancies. The excluded study (Watanabe *et al.*, 2002) was consistent, with a reported median MoM of 0.33.

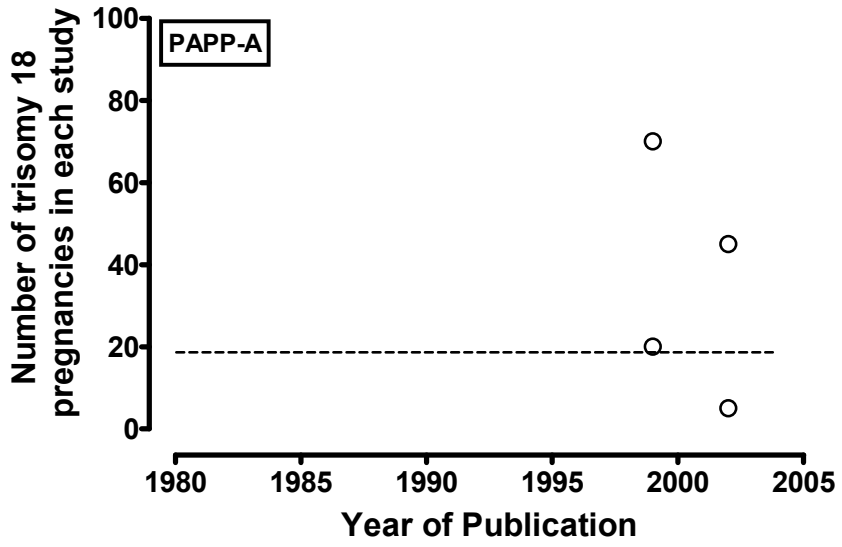

Figure 3.7-1. Publications reporting maternal serum pregnancy associated plasma protein A (PAPP-A) measurements in trisomy 18 pregnancies. The horizontal dashed line is drawn at 19, a reasonable demarcation between small and large studies. Three studies were considered large (between 20 and 70 affected pregnancies) and these were used in the analysis of PAPP-A measurements in trisomy 18 pregnancies.

Table 3.7-1. Maternal serum PAPP-A measurements in trisomy 18 and unaffected pregnancies

Author	GA	Trisomy 18 pregnancies			Unaffected pregnancies		
		N	PAPP-A Central Estimate (MoM)[a]	PAPP-A Log SD	N	Median PAPP-A (MoM)	PAPP-A Log SD
(Muller *et al.*, 2002)	early 2nd	45	-1.0969 (0.08)	0.3700	NR	NR	NR
(Spencer *et al.*, 1999)	14-19	70	-0.9490 (0.11)	0.3840	450	1.00	0.2560
(Bersinger *et al.*, 1999)	15-20	20	-0.8726 (0.11)	0.4408	40	1.35	0.2660
Summary		**135**	**-0.9971 (0.10)**	**0.3994**			
			95% CI (0.08-0.14)				

NR = not reported, GA = gestational age, SD = standard deviation

[a] geometric mean (median). If only one was reported, the other was directly computed.

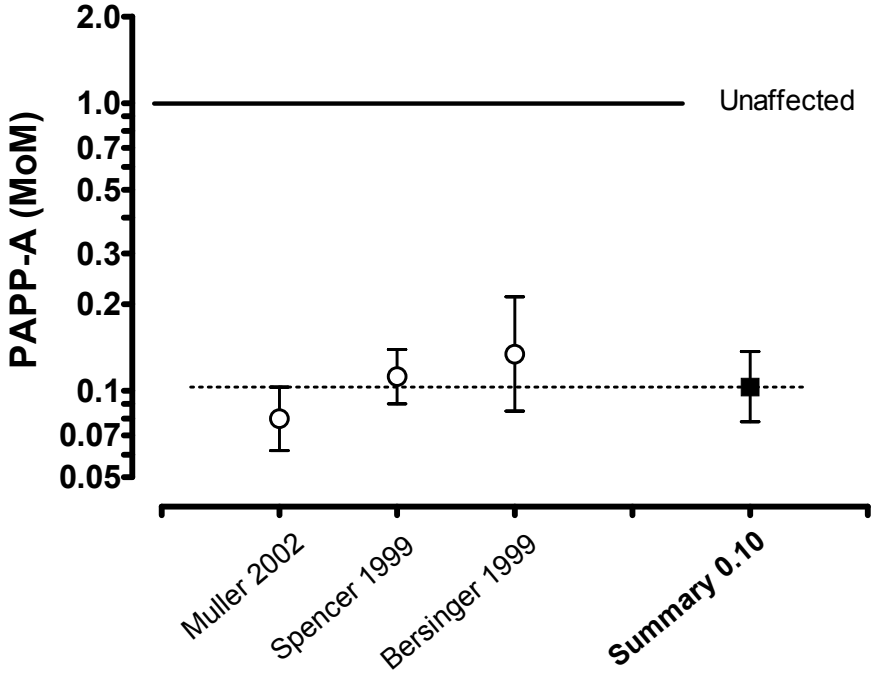

Figure 3.7-2. Maternal serum PAPP-A measurements in trisomy 18 pregnancies. The logarithmic mean and 95% confidence intervals are shown for the three studies reporting PAPP-A measurements from at least 20 pregnancies affected with trisomy 18. The solid line (MoM = 1.00) is the expected value for unaffected pregnancies. The summary logarithmic mean is -0.9871 (PAPP-A median of 0.10 MoM, 95% CI 0.08-0.14).

3.8 Population parameters for trisomy 18 and unaffected pregnancies

This section summarizes the population parameters (logarithmic means, standard deviations) for second trimester maternal serum AFP, uE3, hCG, free β hCG and inhibin-A in trisomy 18 pregnancies reviewed earlier. To model screening performance, it is necessary to also have logarithmic means and standard deviations for unaffected pregnancies. Lastly, pair-wise correlation coefficients for all of these markers in both trisomy 18 and unaffected pregnancies are required as well as truncation limits.

Logarithmic means and standard deviations in trisomy 18 and unaffected pregnancies for six selected second trimester maternal serum analytes

The medians, logarithmic means and standard deviations in trisomy 18 pregnancies in the left-hand side of Table 3.8-1 are taken directly from the summary tables in the preceding sections (3.1 through 3.7). On the right hand side is a recent estimates of the logarithmic standard deviation in unaffected pregnancies for those same serum analytes from the SURUSS study (Wald *et al.*, 2003; Wald, 2006).

Figure 3.8-1 is a visual representation of the overlapping Gaussian curves represented by the parameters in Table 3.8-1, shown as thick solid lines. For comparison, the thinner dashed lines are the trisomy 18 parameters from a commonly used publication (Palomaki *et al.*, 1995). The best marker for trisomy 18 is PAPP-A, while the poorest is inhibin-A.

Table 3.8-1. Summary of logarithmic means and standard deviations (SD) for six second trimester maternal serum analytes in trisomy 18 and unaffected pregnancies

	Trisomy 18 pregnancies			Unaffected pregnancies[a]
Analyte	**Median**	**Log mean[b]**	**Log SD**	**Log SD[b]**
AFP	0.66	-0.1830	0.1817	0.1399
uE3	0.36	-0.4448	0.2817	0.1142
hCG	0.39	-0.4123	0.3561	0.2276
free β hCG	0.24	-0.6203	0.4062	0.2577
Inhibin-A	0.91	-0.0392	0.2897	0.2078
PAPP-A	0.10	-0.9871	0.3894	0.2549

AFP = alpha-fetoprotein, uE3 = unconjugated estriol, hCG = human chorionic gonadotropin, PAPP-A = pregnancy associated plasma protein-A, SD = standard deviation

[a] In unaffected pregnancies, the median is 1.00 and the logarithmic mean is 0.0000. Only the logarithmic SD is reported.

[b] SURUSS – data from N Wald and colleagues (Wald *et al.*, 2003; Wald, 2006)

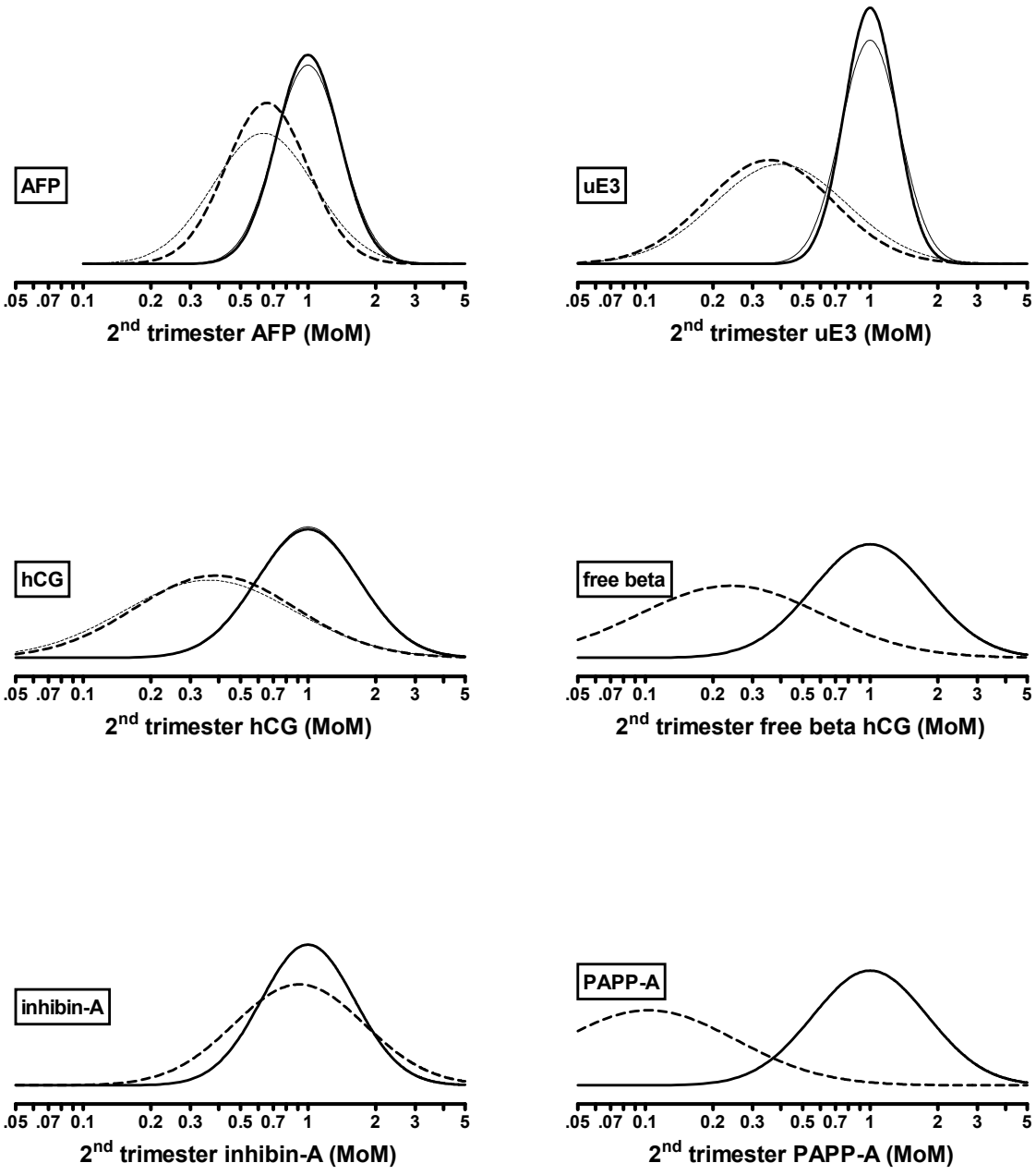

Figure 3.8-1. Overlapping Gaussian curves for six second trimester maternal serum markers in trisomy 18 and unaffected pregnancies. All data are plotted on the same scale. The height of the curves indicates the probability that an observation from that distribution will occur. Data for all six curves are taken from Table 3.8-1 (thick solid and dashed lines). The thin solid and dashed lines indicate the curves currently in used in most programs for second trimester trisomy 18 testing. The dashed line(s) indicates the distributions in trisomy 18 pregnancies, while the solid lines indicate corresponding distributions in unaffected pregnancies.

Pair-wise correlation coefficients in trisomy 18 and unaffected pregnancies

Table 3.8-2 contains the reported correlation coefficients in trisomy 18 and unaffected pregnancies from the studies included in Sections 3.1 to 3.7. The estimates based on at least 40 observations in trisomy 18 pregnancies are bolded. Only three of the pair-wise comparisons are missing. They all include free β hCG measurements (free β hCG versus uE3, hCG and PAPP-A). These are unlikely to be of much consequence to clinical practice. One would normally not measure both the free and intact molecules of hCG at the same time, inhibin-A measurements are not different in trisomy 18 pregnancies so would not be included in any algorithm, and uE3 is not often measured in laboratories that use free β hCG measurements. In general, the correlation coefficients used by most laboratories to generate risk using AFP, uE3 and hCG (Palomaki *et al.*, 1995) measurements are similar to other reported parameters or are centrally located with respect to multiple measurements. For free β hCG versus AFP, two studies (Spencer *et al.*, 1993; Spencer *et al.*, 1999) are consistent in their findings of a small correlation (between 0.12 and 0.20). If PAPP-A measurements are considered the most promising as an addition in the second trimester, the needed correlation coefficients are all available, but from two additional separate studies (Bersinger *et al.*, 1999; Spencer *et al.*, 1999). Because of this 'mix-and-match' for the correlation coefficients from three separate population studies, it will be important for modeling to ensure that computed risks are robust and reliable.

In addition to the correlations in trisomy 18 pregnancies, correlation coefficients from unaffected pregnancies from large published studies are also needed for modeling and assigning trisomy 18 risk. These estimates are contained in the second part of Table 3.8-2, with entries based on at least 1000 observations bolded. Coefficients that are commonly used to assign risk for Down syndrome are also summarized (Knight *et al.*, 1998; Haddow, 1998; Wald *et al.*, 2003). The advantage of the SURUSS parameters (Wald *et al.*, 2003) is that they are comprehensive, but they have the disadvantage of being derived from a relatively small case/control study (<500 samples tested) with measurements occurring over a relatively short time frame.

Table 3.8-2. Correlation coefficients between second trimester serum markers in trisomy 18 and in unaffected pregnancies

	Second trimester serum markers in trisomy 18 pregnancies				
Marker	uE3	hCG	free β hCG	PAPP-A	Inhibin-A
AFP	0.1676 (Lambert-Messerlian *et al.*, 1998) 0.2300 (Sancken *et al.*, 1999) **<u>0.2501</u> (Palomaki *et al.*, 1995)**	-0.1170 (Lambert-Messerlian *et al.*, 1998) **<u>0.0314</u>** (Palomaki *et al.*, 1995) 0.0541 (Kennedy *et al.*, 2000) 0.1251 (Sancken *et al.*, 1999)	**0.1291** (Spencer *et al.*, 1993) **<u>0.2360</u>** (Spencer *et al.*, 1999)	**<u>0.2300</u>** (Spencer *et al.*, 1999)	0.1800 (Lambert-Messerlian *et al.*, 1998)
uE3		-0.2596 (Leporrier *et al.*, 1996) 0.0766 (Lambert-Messerlian *et al.*, 1998) **<u>0.0944</u>** (Palomaki *et al.*, 1995) 0.3500 (Sancken *et al.*, 1999)	NR	-<u>0.0413</u> (Bersinger *et al.*, 1999)	0.4609 (Lambert-Messerlian *et al.*, 1998)
hCG			NR (but 'high')	<u>0.1824</u> (Bersinger *et al.*, 1999)	0.2378 (Lambert-Messerlian *et al.*, 1998)
free β				**0.3400** (Muller *et al.*, 2002) **<u>0.3660</u>** (Spencer *et al.*, 1999)	NR
PAPP-A					0.2100 (Watanabe *et al.*, 2002)

A **bolded entry** indicates that the estimate is based on 40 or more observations.

Underlined estimates will be used in modeling and assigning risk for trisomy 18.

Table 3.8-2. (continued) Correlation coefficients between second trimester serum markers in trisomy 18 and in unaffected pregnancies

	Second trimester serum markers in unaffected pregnancies				
Marker	uE3	hCG	free β hCG	PAPP-A	Inhibin-A
AFP	0.1981 (Wald *et al.*, 2003) **0.2223** (Knight *et al.*, 1998) **0.2610** (Sancken *et al.*, 1999) 0.4737 (Lambert-Messerlian *et al.*, 1998)	**0.1145** (Sancken *et al.*, 1999) -0.1280 (Lambert-Messerlian *et al.*, 1998) 0.1535 (Wald *et al.*, 2003) **0.1560** (Knight *et al.*, 1998) 0.2300 (Aitken *et al.*, 1996)	0.0150 (Spencer *et al.*, 1999) **0.0204** (Spencer *et al.*, 1993) **0.0946** (Knight *et al.*, 1998) 0.0974 (Wald *et al.*, 2003) 0.1500 (Aitken *et al.*, 1996)	-0.0130 (Spencer *et al.*, 1999) 0.1918 (Wald *et al.*, 2003)	-0.0078 (Lambert-Messerlian *et al.*, 1998) 0.2033 (Wald *et al.*, 2003)
uE3		-0.0416 (Wald *et al.*, 2003) **-0.0790** (Sancken *et al.*, 1999) **-0.1400 (Knight *et al.*, 1998)** -0.1614 (Lambert-Messerlian *et al.*, 1998)	-0.0585 (Wald *et al.*, 2003) **-0.1451** (Knight *et al.*, 1998)	0.0983 (Wald *et al.*, 2003) 0.3102 (Bersinger *et al.*, 1999)	-0.1052 (Lambert-Messerlian *et al.*, 1998) -0.0875 (Wald *et al.*, 2003)
hCG			0.8651 (Wald *et al.*, 2003) 0.8700 (Aitken *et al.*, 1996) **0.8757** (Knight *et al.*, 1998)	0.1623 (Bersinger *et al.*, 1999) 0.2838 (Wald *et al.*, 2003)	0.2297 (Lambert-Messerlian *et al.*, 1998) 0.4293 (Wald *et al.*, 2003)
free β				0.0270 (Spencer *et al.*, 1999) 0.2752 (Wald *et al.*, 2003)	0.4092 (Wald *et al.*, 2003)
PAPP-A					0.2530 (Wald *et al.*, 2003)

Bolded entries indicate that the estimate is based on 1,000 or more observations.

Estimates from SURUSS (Wald *et al.*, 2003) are underlined, and will be used for modeling and assigning risk.

3.9 Modeling performance of serum markers

The truncation limits for each of the markers also need to be specified. These are already available for AFP, uE3 and hCG measurements. Table 3.9-1 provides reasonable truncation limits for Free β and PAPP-A measurements that are consistent with the methodology described in Chapter 3, Section 3.1. These truncation limits can also be examined in relation to the overlapping curves shown in Figure 3.8-1. At the upper and lower truncation limits, the height of the curves in unaffected and trisomy 18 pregnancies are above baseline, indicating that these limits are not in the extreme tails of either of the distributions.

Table 3.9-1. Truncation limits (TL) for maternal serum markers

	(Palomaki *et al.*, 2005)		Recommended	
Marker	**Lower TL**	**Upper TL**	**Lower TL**	**Upper TL**
AFP	0.33	2.00		
uE3	0.40	1.50		
hCG	0.20	2.50		
Free β	NR	NR	0.20	2.50
PAPP-A	NR	NR	0.20	1.00

The set of population parameters listed in the previous Sections (Tables 3.8-1, 3.8-2 and 3.9-1), can now be combined with the age-associated risk for trisomy 18 to assign individual patient-specific risks. By including the known distribution of maternal ages in a defined population, it is possible to construct a *monte carlo* simulation to model the trisomy 18 detection rate and associated false positive rates using trisomy 18 risk as the test result. The distribution of maternal ages in England and Wales for 2006 through 2008 can be used as a good approximation of the distribution of maternal ages in unaffected pregnancies, as the prevalence of age-associated disorders (*e.g.,* Down syndrome and trisomy 18) are relatively rare. Using a published equation to assign age-related risk for trisomy 18 (Savva *et al.*, 2010), one can also estimate the distribution of maternal ages for women with a trisomy 18 fetus at term. These two overlapping distributions are shown in Figure 3.9-1. Using these numbers, the overall birth prevalence is 2.9/10,000.

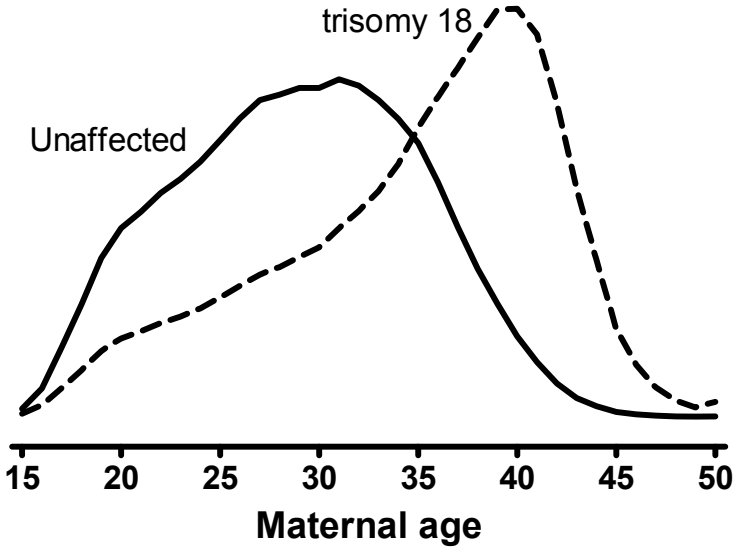

Figure 3.9-1. Maternal age distribution in unaffected and trisomy 18 births. The solid line is the observed distribution of maternal ages in England and Wales in 2006 through 2008, derived from published data (Morris and Savva, 2008). The dashed line is the expected distribution of trisomy 18 births in the absence of prenatal diagnosis and selective termination.

Table 3.9-2 shows the results of modeling for six combinations of maternal age and serum markers. Specifically, double markers (AFP and hCG, or AFP and free β), triple markers (double plus uE3) and quadruple markers (triple plus PAPP-A). The table is further divided into two groups of rows. The first group shows the false positive rates for these combinations at a fixed detection rate, while the second group shows detection rates at fixed false positive rates. As the number of markers increases, overall test performance increases considerably. For example, at a fixed detection rate of 80%, the inclusion of uE3 reduces the false positive rate from 9.0% to 0.3%, with a further reduction to <0.1% with the addition of PAPP-A measurements. Replacing hCG measurements with those of the free β subunit improves performance a great deal for the double markers (9.0% versus 5.3% false positive rate for an 80% detection rate), but less so for triple and quadruple marker testing (0.3% versus 0.2% and <0.1% versus <0.1%, respectively).

In Table 3.9-3, the modeling results for the same combinations are shown by risk cut-off level and include not only the detection and false positive rates, but also the odds of being affected given a positive result (OAPR). Term risks are used to allow for

comparisons between these second trimester risks and those first trimester algorithms that will be evaluated later. Performance for the double markers is limited, but reasonable cut-off levels for the triple and quadruple marker combinations might be term risks of 1:300 (second trimester risk of about 1:100). The corresponding second trimester OAPRs are both very high at 1:5 and 1:2, respectively. The major difference between the three and four marker algorithms is a three-fold reduction in the false positive rate (0.39% to 0.11%) with a concomitant improvement in detection rate (from 81% to 88%. These very low false positive rates may be somewhat misleading. Relatively rare events (*e.g.*, existing fetal deaths, anencephaly) might also be preferentially assigned high risks, but would not be considered 'false positives' in the usual sense of the term. Actual clinical performance will likely be associated with an additional two or three per 1000 more positive test results that those predicted based on modeling alone.

A direct comparison can also be made between these modeling results and those reported in the literature (Palomaki *et al.*, 1995) and summarized in Table 3.11-1. Two clear points emerge. First, the test performance is better. Using triple marker testing with AFP, uE3 and hCG as an example, the 1995 modeling estimated a detection rate of 60% at a false positive rate of 0.2%. The current modeling estimates 80% detection at a false positive rate of 0.3%. This improvement can be traced to two factors: the standard deviation of AFP measurements in affected pregnancies is considerably tighter than expected, and the standard deviation of uE3 measurements in unaffected pregnancies is also considerably tighter. Some improvement may also be attributed to the underlying maternal age distribution used being considerably older than that used in 1995. All of these factors will improve screening performance.

The second point is that select performance occurs at quite different risk cut-off levels. For example, using the double test with AFP and hCG, the 1995 modeling estimated using a 1:100 second trimester risk cut-off level, the detection rate would be 30% and the false positive rate would be 0.2%. This is similar to the performance seen for the current modeling (29% detection at 0.1% false positive rate from Table 3.9.3). However, this performance estimate occurs using a second trimester risk cut-off of 1:18. Although some of this could be attributed to improved performance, most is due to recent updates in the age-associated prior risk of trisomy 18 as well as the second trimester fetal loss rate.

Table 3.9-2. Modeled trisomy 18 detection rates (DR) and false positive rates (FPR) using second trimester maternal serum markers

	Maternal age and AFP, in combination with screening markers					
	hCG	**free β**	**uE3 & hCG**	**uE3 & free β**	**uE3, hCG & PAPP-A**	**uE3, free β & PAPP-A**
DR (%)	**False Positive Rate (%)**					
50	0.7	0.3	<0.1	<0.1	<0.1	<0.1
60	1.6	0.8	<0.1	<0.1	<0.1	<0.1
70	3.8	2.0	<0.1	<0.1	<0.1	<0.1
80	9.0	5.3	0.3	0.2	<0.1	<0.1
90	>20	16	2.9	1.1	0.2	0.1
FPR (%)	**Detection Rate (%)**					
0.3	39	48	79	83	92	94
0.5	45	54	82	86	94	96
0.7	49	58	83	88	96	97
1.0	54	62	84	90	97	97
1.5	59	67	87	91	98	98

Table 3.9-3. Modeled trisomy 18 detection rates (DR), false positive rates (FPR) and odds of being affected given a positive test result (OAPR) using second trimester maternal serum markers at selected trisomy 18 risk cut-off levels

	Maternal age in combination with screening markers								
	Double Test			Triple test			Quadruple test		
Risk	AFP & hCG			AFP, uE3 & hCG			AFP, uE3, hCG & PAPP-A		
at term (2^{nd})[1]	DR	FPR	OAPR[2]	DR (%)	FPR (%)	OAPR[2]	DR (%)	FPR (%)	OAPR[2]
1: 50 (1: 18)	29	0.10	1: 5	68	<0.10	-	76	<0.10	-
1:100 (1: 35)	38	0.27	1: 9	73	0.12	-	81	<0.10	-
1:150 (1: 53)	44	0.41	1:12	76	0.18	1:3	84	<0.10	-
1:200 (1: 70)	48	0.63	1:17	78	0.26	1:4	85	<0.10	-
1:250 (1: 88)	51	0.83	1:21	80	0.33	1:5	87	<0.10	-
1:300 (1:105)	54	1.0	1:24	81	0.39	1:6	88	0.11	1:2
1:350 (1:123)	57	1.2	1:27	81	0.45	1:7	88	0.12	1:2
1:400 (1:140)	59	1.5	1:33	82	0.51	1:8	89	0.13	1:2

[1] assuming a 65% fetal loss from 16-18 weeks to term (Table 2.4-1)

[2] OAPR = second trimester odds of being affected given a positive result (assumes birth prevalence of 2.69/10,000, with adjustment for 65% fetal loss from 16-18 weeks to term).

Figure 3.9-2 shows the age-associated risk of trisomy 18, based on one-tenth the age-associated term risk of Down syndrome (Hecht and Hook, 1996) as well as the risk based on a recent publication using direct observations of trisomy 18 birth pregnancies (Morris 2008). At 25 and 40 years of age, for example, the current risk estimates (solid line) are higher by a factor of 1.6 (1:13,100 versus 8,350) and 2.3 (1:994 versus 1:435), respectively. In 1995, these term risks were converted to second trimester using a fetal loss rate of 70% compared to the current estimate of 65%, increasing the difference even further. Lastly, the general shifting towards older pregnancies will also cause the risk cutoff levels to shift towards higher values.

However, the reasonable risk cutoff-level for both the old and new modeling for the triple test is a second trimester risk of about 1:100, easing the implementation of the new model. The suggested changes in the underlying statistical parameters would not be noticeable to health providers, but the performance would improve, especially if PAPP-A measurements were to be included.

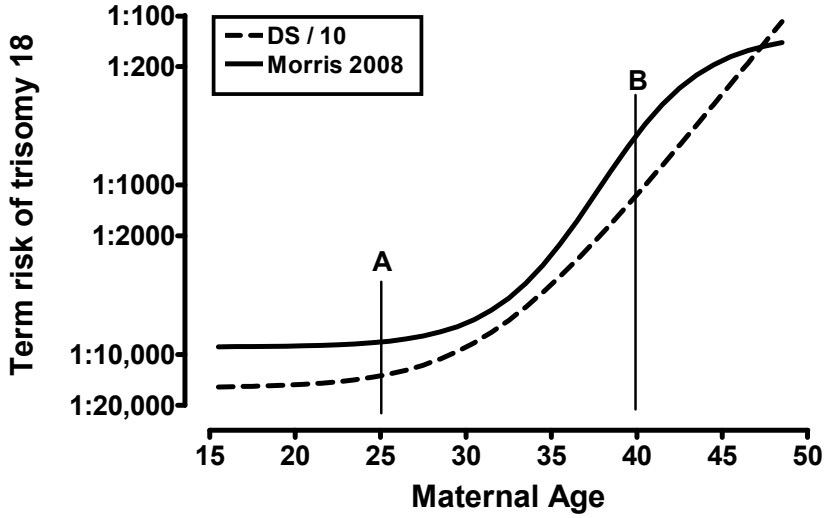

Figure 3.9-2. A comparison of two methods of assigning age-associated term risks for trisomy 18. The solid line (Morris and Savva, 2008) shows the modeled age-specific trisomy 18 risk at term. Prior to this, an alternative method was used to approximate trisomy 18 risk. This was done by dividing the Down syndrome term risk (Hecht and Hook, 1996) by 10. The factor 10 was the approximate ratio of term Down syndrome births to trisomy 18 births. At 25 (A) and 40 years (B) of age, the corresponding risks are 1:8,350, 1:13,100 and 1:994, 1:435, respectively.

3.10 Demonstration studies using fixed MoM cut-offs

In the 1980s, when only maternal age and AFP measurements were being used to screen for Down syndrome in the second trimester, there was no need for a separate algorithm for identifying pregnancies at high risk for trisomy 18. This is because both Down syndrome and trisomy 18 pregnancies have lower AFP measurements, and both aneuploidies are associated with increasing maternal age. If a woman is at increased risk of Down syndrome based on age and AFP measurements, then she would also be at increased risk of trisomy 18. The information about increased risk for trisomy 18 was often provided as part of counseling women screen positive for Down syndrome, prior to diagnostic testing. Two studies in the late 1980s included information about the identification of trisomy 18 pregnancies as part of a maternal age/AFP screening program for Down syndrome. The Yale program screened 24,065 women under age 35 for Down syndrome, with 6.0% of the population determined to be at or above the risk of a 35 year old woman (DiMaio *et al.*, 1987). Among these screen positive pregnancies, three were identified as having trisomy 18 at the time of amniocentesis. A second multicenter trial in New England (Palomaki, 1989) enrolled 77,273 women under age 35 from eight centers. A total of 4.7% were screen positive for Down syndrome and four cases of trisomy 18 were identified. Neither group attempted to determine a trisomy 18 detection rate.

At about the same time that these Down syndrome demonstration studies were being reported, new information about additional second trimester serum markers was emerging, including uE3, hCG and the free β subunit of hCG. As documented earlier, some studies of these newer markers included reports of trisomy 18, and it soon became clear that the pattern of low, low, high (for AFP, uE3 and hCG) found in Down syndrome was not the same as the pattern seen in trisomy 18 pregnancies. AFP and uE3 measurements were still both low in trisomy 18 pregnancies, but hCG measurements were also low, not high. Because of this important difference, double or triple marker algorithms for Down syndrome were less likely to also identify pregnancies with trisomy 18. A new algorithm targeted towards trisomy 18 was needed. In 1990, Canick and his colleagues (Canick *et al.*, 1990) published the levels of AFP, uE3 and hCG in a series of 10 trisomy 18 pregnancies and proposed a simple algorithm based on reduced measurements of all three analytes. Using fixed MoM cut-off levels of $\leq$0.75 MoM, $\leq$0.60 MoM and $\leq$0.55 MoM for AFP, uE3 and hCG, respectively, they estimated that 60% of trisomy 18 pregnancies might be identifiable by offering amniocentesis to about 0.4% of

women. In the second trimester, this would result in an odds of being affected given a positive result (OAPR) of about 1:16 (alternatively, a positive predictive value of 5.9%). Some argued that the fixed cut-off levels should be set higher, so that a higher detection rate could be achieved. However, this would result in a higher false positive rate and, therefore, a lower OAPR. The OAPR is directly translated into the number of procedures needed to identify a case of trisomy 18 in the second trimester. Given the relatively low chance of survival to term, and the high rate of death among the few live births, it was decided to set the fixed MoM cut-offs to maintain a high target OAPR.

The fixed MoM cut-off algorithm for the triple markers AFP, uE3 and hCG was widely adopted in the United States and used extensively in the 1990s. The majority of studies reported trisomy 18 results as part of a Down syndrome demonstration study, but several focused only on trisomy 18 testing. Because this algorithm is no longer in use, only summary information will be provided. A total of 10 studies used AFP, uE3 and hCG (Palomaki *et al.*, 1992; Burton *et al.*, 1993; Bradley, 1994; Kellner *et al.*, 1995; Benn *et al.*, 1996; McDuffie *et al.*, 1996; Yankowitz *et al.*, 1998; Feuchtbaum *et al.*, 2000; Hogge *et al.*, 2001; Summers *et al.*, 2003) and 2 used AFP and free β hCG (Seppo *et al.*, 1999; Chao *et al.*, 1999). A total of 316,655 women were tested, and 0.29% (95% CI 0.21-0.41%) had the high risk triple marker pattern. Among these 1,639 women, 99 trisomy 18 fetuses were identified for an OAPR of 1:15 (95% CI 1:10 to 1:26).

3.11 Demonstration studies using trisomy 18 risk

The fixed MoM cut-off model does not account for the known age-associated change in the prior risk of trisomy 18. More importantly, the results of the interpretation are dichotomous (positive, negative) and, therefore, cannot differentiate between those pregnancies that have markers near the cut-off levels or much below those levels. Clinicians and those involved in screening programs needed to be able to assign patient-specific risks for trisomy 18 in much the same way as when screening for Down syndrome. However, such a model would require a relatively large number of unbiased cases with AFP, uE3 and hCG measurements. The cases would also need to have been identified without regards to the biochemistry results. For example, stored serum samples collected prior to implementation of the fixed MoM cut-off trisomy 18 protocol from women whose pregnancy was later found to have trisomy 18. Diagnostic testing might have been triggered by abnormal ultrasound findings or by clinical finding at the time of birth. Other sources of unbiased samples would be those collected prior to amniocentesis performed due to maternal age or family history of aneuploidy.

In 1995, a collaborative study of 94 second trimester serum markers in trisomy 18 fetuses was published by our research group (Palomaki *et al.*, 1995). Of these, 89 were from pregnancies without an open defect. All pregnancies had existing triple marker testing results. A risk algorithm was developed, based on the same concept of overlapping Gaussian distributions as was used for assigning Down syndrome risk. An important advantage of this dataset was the range of indications for diagnosis. In 38%, the samples were collected due to age 35 or older, in another 17%, the indication was abnormal ultrasound, another 13% were live births, with the remaining 31% having only AFP screening. This latter category was only included if the center was able to identify live-born/stillborn cases as well. By comparing the serum marker measurements among these four groups, it was possible to determine whether any group differed in any important way from another. Overall, no differences were found for the AFP or uE3 measurements. However, the group of older women did tend to have lower than expected hCG measurements (0.26 MoM vs. 0.42 MoM, p=0.05).

Modeling based on these published parameters suggested that using the risk algorithm with triple markers would provide considerable improvement by both increasing the detection rate and reducing the false positive rate. Table 3.11-1 summarizes the modeling results from that publication (Palomaki *et al.*, 1995). This algorithm is currently the standard of care for assigning trisomy 18 patient-specific risks in the second

trimester, with the boxed row indicating the most commonly used second trimester risk cut-off level of 1:100. From this table, it is clear that a double marker algorithm (AFP and hCG, or AFP and the free β subunit of hCG) is less effective in identifying trisomy 18 pregnancies than the triple test.

The English literature was searched as described earlier. Several of the studies (Benn *et al.*, 1996; Summers *et al.*, 2003) reported results for both the fixed MoM cut-off and risk protocols. The majority of studies reported the trisomy 18 screening results as part of a Down syndrome demonstration study, but several focused only on trisomy 18. All used patient-specific risks based on maternal age, AFP, uE3 and hCG measurements and used a published algorithm (Palomaki *et al.*, 1995). Figure 3.11-1 shows the six studies included in the analyses versus the year of publication (Benn *et al.*, 1999; Hogge *et al.*, 2001; Summers *et al.*, 2003; Jaques *et al.*, 2007; Breathnach *et al.*, 2007; Wortelboer *et al.*, 2008). Since many laboratories were using fixed MoM cut-off levels routinely, it took several years for programs to switch methodologies, implement testing, and report results. In some of the publications, the risks were assigned retrospectively, for samples collected prior to 1995 (Benn *et al.*, 1999; Hogge *et al.*, 2001).

Table 3.11-1. Test performance for trisomy 18 in the second trimester using maternal age and measurements of maternal serum AFP, uE3 and hCG to assign patient-specific risks

Term (2nd trim)	Maternal age in combination with					
Risk Cut-off	AFP and hCG			AFP, uE3 and hCG		
(≥ 1:n)	DR (%)	FPR (%)	OAPR	DR (%)	FPR (%)	OAPR
1: 170 (1: 50)	22	0.1	1:11	33	<0.1	-
1: 330 (1:100)	30	0.2	1:16	60	0.2	1: 8
1: 500 (1:150)	38	0.5	1:32	65	0.3	1:12
1: 670 (1:200)	43	0.7	1:39	68	0.4	1:14
1:1000 (1:300)	49	1.2	1:59	70	0.6	1:21
1:1300 (1:400)	54	1.7	1:76	73	0.8	1:26
1:1700 (1:500)	57	2.3	1:97	76	1.0	1:32

AFP = alpha-fetoprotein, uE3 = unconjugated estriol, hCG = human chorionic gonadotropin
DR = detection rate, FPR = false positive rate, OAPR = odds of being affected given a positive result

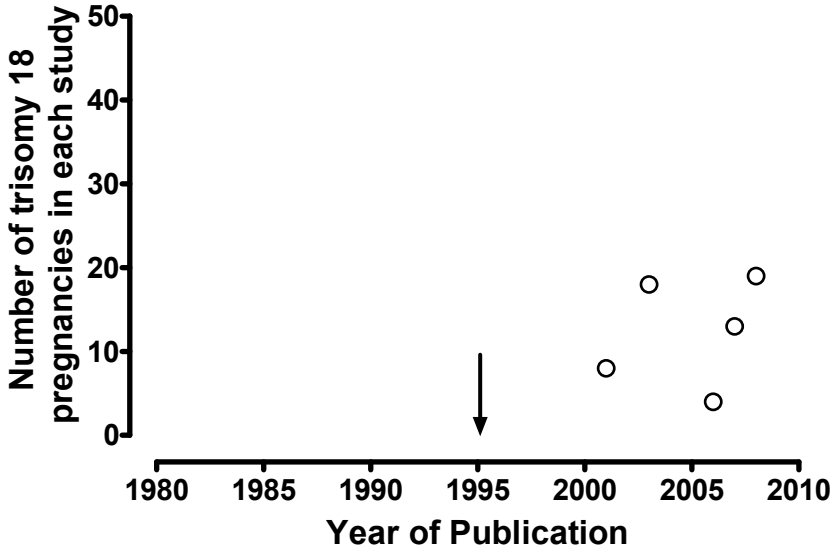

Figure 3.11-1. Publications reporting the results of second trimester trisomy 18 demonstration studies that used patient-specific trisomy 18 risks. The risk algorithm used in these reports was first published in 1995 (arrow). All reported trials use maternal serum AFP, uE3 and hCG measurements. None of the studies reported 20 or more cases of trisomy 18 detected.

Table 3.11-2 shows summary information from the six included studies. Overall, the summary odds of being affected given a positive result (OAPR) is 1:14 (95% CI 1:8 to 1:23) with considerable heterogeneity (Q=12, I^2=66%, p=0.02). This is somewhat lower than the predicted OAPR for this algorithm of 1:8 in Table 3.10-2 (Palomaki *et al.*, 1995). There may be a reasonable explanation for the lower than expected OAPR. The algorithm not only identifies trisomy 18, but also existing fetal deaths, anencephaly and other abnormalities (Palomaki *et al.*, 1995; Benn *et al.*, 1996). Two of the intervention trials (Benn *et al.*, 1996; Hogge *et al.*, 2001) also reported the number of fetal deaths/abnormalities, and the revised positive rates are considerably lower (0.47% to 0.35% and 0.55% to 0.41%, respectively). These two studies had the lowest OAPR estimates of 1:21.6 and 1:31.3, and after removing fetal death/abnormalities, these were increased to 1:16 and 1:22, respectively. A higher than expected number of 'false positive' are likely with the very specific trisomy 18 algorithms due to other true abnormal (but not trisomy 18) outcomes.

Table 3.11-2. Trisomy 18 demonstration studies using patient specific risk

Study	Years Included	Number	Age 35 or older (%)	Screen Positive Number	Screen Positive (%)	Screen Positive Revised[1]	Number of T18	OAPR (1:n)
(Benn *et al.*, 1999)	92 to 97	41,565	NR	194	0.47	147	9	1:21.6
(Hogge *et al.*, 2001)	93 to 98	45,145	10.2	250	0.55	183	8	1:31.3
(Summers *et al.*, 2003)	98 to 00	107,240	16.2	223	0.21	NR	18	1:12.4
(Jaques *et al.*, 2007)	98 to 00	16,607	NR	41	0.25	NR	4	1:10.3
(Breathnach *et al.*, 2007)	99 to 02	35,120	NR	109	0.31	NR	13	1: 8.4
All		**245,677**		**817**	**0.33**		**52**	**1:14**
				(95% CI 0.22 – 0.51)				(95% CI 1:8 – 1:23)
(Wortelboer *et al.*, 2008)[2]	91 to 05	42,554	NR	NR	NR	NR	19	NR

OAPR = odds of being affected given a positive result (equivalent to the positive predictive value)

[1] Number screen positive after other known abnormalities and existing fetal demises removed.

[2] Study not included in the summary analysis (All), due to limited data.

The analysis was performed on the logit of the PPV [logit (PPV) = ln(PPV/(1-PPV))]. The summary estimate based on the random effects modeling of these logit values was then converted back to an odds of being affected given a positive result (OAPR). For example, the summary logit was -2.618. The corresponding OAPR is 1:14 and is computed as follows:

PPV (as a risk)	=	exp(-2.618) / (1+exp(-2.618))
PPV (risk)	=	0.07295 / (1.07295) = 0.06799
PPV (as an odds)	=	1:(1 - 0.06799)/(0.06799) = 1:14

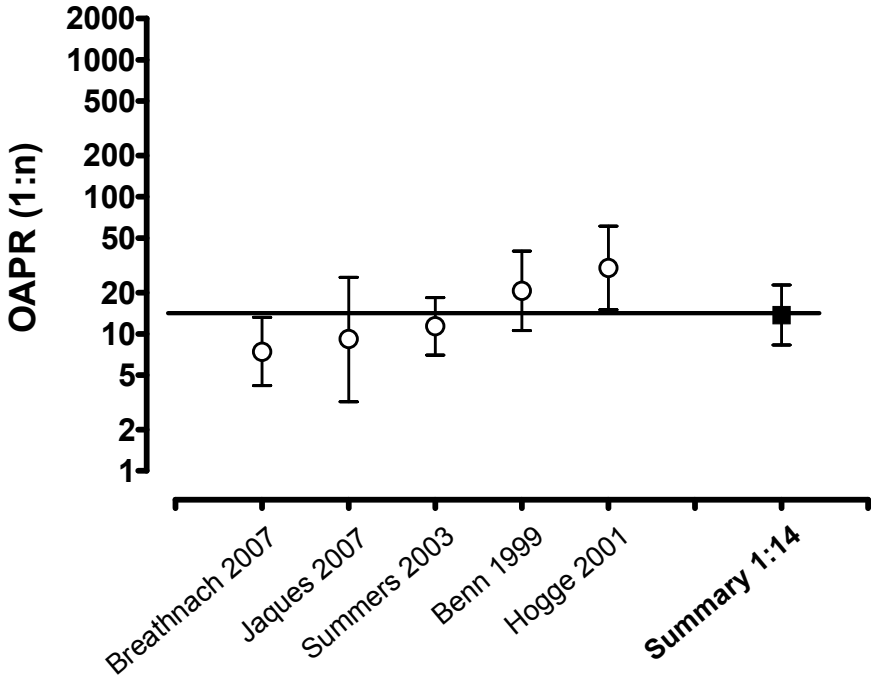

Figure 3.11-2. The odds of being affected given a positive result (OAPR) for trisomy 18 screening trials using patient-specific risks. These studies assigned risk using AFP, uE3 and hCG measurements. The logit of the OAPR is analyzed using a random effects model, with the summary OAPR of 1:14 (logit of -2.618).

3.12 Conclusions

Five questions were posed at the beginning of this section, and they can now be answered.

Q1. Based on observational studies, what are the summary population parameters for maternal serum markers in second trimester trisomy 18 pregnancies, and do they agree with those in wide use today (Palomaki *et al.*, 1995)?

Reasonable parameters are summarized in tables and figures contained in Section 3.8. The AFP, uE3 and hCG parameters derived from the literature are close to those used routinely for assigning trisomy 18 risk in clinical practice (Palomaki *et al.*, 1995), with the exception of a modest tightening of the AFP logarithmic standard deviation (from 0.2239 to 0.1817). Although it is not necessary to revise the parameters for AFP, uE3 and hCG for trisomy 18, the updated population parameters could be used in clinical practice and would likely result in more appropriate patient-specific risks. Programs could also take this opportunity to update the age-specific term risks for trisomy 18, as well as the fetal loss rates from the early second trimester to term.

Q2. Do trisomy 18 demonstration studies using AFP, uE3 and hCG confirm the performance of the risk-based model (Palomaki *et al.*, 1995), as defined by the false positive rate and positive predictive value?

It is difficult to confirm a detection rate for trisomy 18 as part of a demonstration study, but the reported false positive rates and odds of being affected given a positive result (OAPR) meet expectations. For example, the expected false positive rate for the risk-based algorithm was 0.2%, and the observed summary rate was 0.33 (95% CI 0.22 – 0.51). The expected OAPR was 1:8, and the observed rate was 1:14 (95% CI 8 to 23). Both of these rates are somewhat short of target, because this algorithm also preferentially identifies other abnormal pregnancies outcomes besides trisomy 18 (Section 3.10).

Q3. Are measurements of inhibin-A useful to add to the risk-based model? If so, describe the algorithm and model the expected increase in performance.

Based on the two largest studies, the reduction in median inhibin-A levels in trisomy 18 is only about 9% (0.91 MoM) and is not statistically significant. Other smaller published studies report that the levels were 'not different', or were consistent with little or no change. Inhibin-A does not appear to be useful for trisomy 18 testing in the second trimester.

Q4. Are there other serum markers that may be of use in a risk-based model? If so, describe the algorithm and model the expected increase in performance.

PAPP-A measurements are extremely low in trisomy 18 pregnancies in the early second trimester. Were these measurements to be routinely available, performance would be significantly enhanced. For example, at a detection rate of 80%, the false positive rate would drop from 1.5% to about 0.1%. Screening programs would need to evaluate the costs of adding a second trimester PAPP-A against improved detection, reduced false positives, or a combination of the two.

Although a complete cost benefit analysis is beyond the scope of this project, with some simplifying assumptions, one can determine whether more extensive analyses are warranted. If the trisomy 18 detection rate were held constant at 80%, the expected false positive rate would be reduced from 3/1000 to 1/1000 (Table 3.9.2) by adding PAPP-A measurements. Among a population of 10,000 pregnancies tested, this translates into detecting about 7 of the 8 cases occurring, while identifying 30 false positive results using the triple test. Adding PAPP-A reduces this to compared to 10 false positives with a savings of 20 amniocenteses / karyotypes. A a cost of about $1000 each, this represents a $20,000 savings and translates into $2 per patient available for the PAPP-A measurement. This is insufficient to cover the $5 or so needed to run a PAPP-A test. However, contingent testing can greatly reduce the costs associated with testing everyone in the population while closely maintaining performance (Palomaki et al., 2006). This is especially true when the marker is highly predictive, such as is in this scenario with PAPP-A. Contingent models that include PAPP-A measurement on even one-quarter (or less) of the population would then be able to reduce the overall health care costs. At the same time, fewer procedure-related losses would occur due to the 67% reduction in invasive procedures. Given that many laboratories have PAPP-A assays available for first trimester testing, this is a real potential improvement in second trimester testing for trisomy 18.