Repeated measures screening for down's syndrome

Bradbury, I., & Wright, D. E. (2005). Repeated measures screening for down's syndrome. *British Journal of Obstetrics and Gynaecology*, 112 (1)(1), 80-83. <https://doi.org/10.1111/j.1471-0528.2004.00343.x>

Published in:
British Journal of Obstetrics and Gynaecology

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Repeated measures screening for Down's Syndrome

David E. Wright,^a Ian Bradbury^b

Objective To demonstrate the potential value of screening for Down's Syndrome using highly correlated repeated measures of serum markers taken in the first and second trimesters of pregnancy.

Design A Monte Carlo simulation study.

Population Detection rates and false positive rates relating to the maternal age distribution of England and Wales for the period 1996 to 1998 were obtained using marker distributions from the SURUSS study.

Results Screening using first trimester nuchal translucency and repeated measures of uE3 and PAPP-A in the first and second trimester has an estimated false positive rate of 0.3% for an 85% detection rate. This should be compared with the integrated test with an estimated false positive rate of 1.2% for the same detection rate.

Conclusions The performance of repeated measures screening tests, and their acceptability to women, should be assessed in further prospective studies.

INTRODUCTION

Nowadays, multivariate combinations of markers are used routinely in the first and second trimesters of pregnancy. The recently introduced integrated test,¹ combines measures of nuchal translucency (NT) and PAPP-A from the first trimester with measures of α -fetoprotein, uE3, free β -hCG and inhibin-A in the second trimester. Modelling the performance of this test on the population of maternal ages for England and Wales over the period 1996 to 1998, Wald *et al.*² report a false positive rate of 1.2% for an 85% detection rate. This should be compared with the quadruple test (α -fetoprotein, uE3, free β -hCG and Inhibin-A) in the second trimester and the combined test (NT, free β hCG and PAPP-A) in the first trimester which have false positive rates of 6.2% and 6.1%, respectively, for the same 85% detection rate. As Wald *et al.*² discuss, the lower false positive rate associated with the integrated test has substantial advantages in financial and human terms resulting from the reduction in invasive follow up tests.

The choice of markers in multimer screening tests has been influenced by the extent to which they provide 'independent information' as characterised by low correlations between markers and the univariate properties of markers. The prevailing view has been that combining markers with low correlations that individually have good discriminatory power represents the best approach to screening test design. The integrated test² was obtained by combining the best

markers in the first and second trimesters. In this paper, we demonstrate that certain combinations of highly correlated markers, some of which individually have poor discriminatory power, have substantial benefits over the established combinations of markers used in the integrated test.

METHODS

Estimates of means, standard deviations and correlations of log multiples of medians were obtained from the SURUSS study.² This study involved 47,053 pregnancies including 101 cases of Down's Syndrome. The maternal age related risk of Down's Syndrome was obtained from the model of Wright and Bray.³ Detection rates and false positive rates were estimated for the maternal age distribution of England and Wales for the period 1996 to 1998⁴ using the methodology described below.

Detection rates and false positive rates were obtained using Monte Carlo methods to sample from the joint distribution of log multiples of medians in Down's Syndrome and in unaffected pregnancies. The risk cutoff above which pregnancies were screened positive was determined to fix either the false positive rate or detection rate. This enables different screening tests to be compared in terms of false positive rate or detection rate for a fixed detection rate or false positive rate, respectively. Samples of 500,000 observations were drawn from the multivariate Gaussian distribution for Down's Syndrome and unaffected pregnancies. Likelihood ratios were computed for each observation and then used to compute detection rates and false positive rates for each maternal age. Overall population detection rates were obtained by combining maternal age specific rates according to the maternal age distribution for Down's Syndrome. Similarly, false positive rates were obtained using the maternal age distribution of unaffected pregnancies. The precision of the

^aSchool of Mathematics and Statistics, University of Plymouth, UK

^bSchool of Biomedical Science, University of Ulster, Coleraine, UK

Correspondence: Dr D. E. Wright, School of Mathematics and Statistics, University of Plymouth, Plymouth, PL4 8AA UK.

estimated detection rates and false positive rates was assessed both theoretically and empirically and 500,000 provide sufficiently precise estimates of detection and false positive rates. Of course, this accounts only for the imprecision due to the Monte Carlo approach and not for the variability of the estimated means, standard deviations and correlations used in the calculations. We remark that, apart from the model for the maternal age risk, our assumptions are the same as those in the SURUSS study²; to the number of decimal places presented here, the detection rates and false positive rates are identical to those obtained from the maternal age risk model used in the SURUSS study.² Documentation and copies of the S-Plus functions used in the estimation of detection rates and false positive rates are available on request from the first author.

It is well known that the methodology described above is optimistically biased in the sense that it will produce better screening performance measures (lower false positive rates for a fixed detection rate) than would result from a prospective evaluation.⁵ This is because the parameter estimates used in the algorithm are subject to sampling error and will differ from those in the population. By assuming the same

parameter estimated for the population and algorithm, the screening algorithm is optimised in a way that it would not be in practice.

RESULTS

Table 1 gives a summary of the screening performance of various tests using the methodology described above. The notation $\wedge^2$ is used to represent repeated measures of a particular marker in the first and second trimester, so for example, PAPP-A $\wedge^2$ denotes first and second trimester PAPP-A.

Some of the results presented in Table 1 run counter to the established view on the use of markers in screening. In particular, repeated measures of PAPP-A are highly correlated (0.8798 and 0.6895 in unaffected and Down's Syndrome pregnancies, respectively). Moreover, when viewed in isolation, PAPP-A in the second trimester is a very poor marker for Down's Syndrome. Yet the addition of second trimester PAPP-A to first trimester PAPP-A reduces the false positive rate for an 85% detection rate from 16% to 2.3%. The mechanism by which the addition of the second trimester measure of PAPP-A improves screening performance is illustrated in Fig. 1. This shows samples of $\log(\text{MoM PAPP-A})$ in the first trimester (horizontal axis) and the second trimester (vertical axis) for 1000 Down's Syndrome pregnancies and 1000 unaffected pregnancies. These samples were drawn from bivariate Gaussian distributions using parameter estimates taken from the SURUSS study.² The ellipses represent contours containing 90% of the distribution of values. Figure 1 demonstrates that even though the individual markers do not provide good discrimination, the *joint* distribution of the two is effective in separating the two populations. Contrary to the intuition that highly correlated markers are unlikely to be useful, joint discrimination of this type is more likely to occur when the two marker values (or values of one marker in the two trimesters) show high correlation.

Inspection of Table 1 shows that combinations of PAPP-A $\wedge^2$ with uE3 $\wedge^2$ and of PAPP-A $\wedge^2$ with total hCG $\wedge^2$ give particularly good discrimination. The combination of PAPP-A $\wedge^2$, Inhibin $\wedge^2$ and uE3 $\wedge^2$ also has a relatively low false positive rate. More detailed performance characteristics of these tests are shown in Table 2. This demonstrates substantial improvement in screening performance over the integrated tests obtained by combining the best first trimester markers and the best second trimester markers in terms of safety, cost and detection.

DISCUSSION

The results presented here are based on multivariate Gaussian models for the distribution of markers and they fail to take account of uncertainty in the parameter estimates. This uncertainty is likely to be particularly important

Table 1. False positive rates for an 85% detection rate using repeated measures screening tests. In the repeated measures test, it is assumed that first trimester sampling takes place at 10 weeks of gestation. The detection rates and false positive rates relate to the maternal age distribution of England and Wales for the three years 1996 to 1998. It is assumed that MoM values are obtained from scan estimates of gestational age and are corrected for weight.

	Markers included	False positive rate (%)
First trimester combined test	NT, free β -hCG, PAPP-A	6.1
Second trimester quadruple test	α -Fetoprotein (AFP), uE3, hCG, inhibin	6.2
Two measures	AFP $\wedge^2$	41
	UE3 $\wedge^2$	17
	Total hCG $\wedge^2$	13
	Free β -hCG	21
	PAPP-A $\wedge^2$	2.3
	Inhibin-A $\wedge^2$	12
Four measures	Total hCG $\wedge^2$ + PAPP-A $\wedge^2$	1.1
	Total hCG $\wedge^2$ + inhibin-A $\wedge^2$	5.7
	Total hCG $\wedge^2$ + uE3	5.3
	PAPP-A $\wedge^2$ + inhibin-A $\wedge^2$	1.0
	PAPP-A $\wedge^2$ + uE3 $\wedge^2$	0.5
Six measures	PAPP-A $\wedge^2$ + total hCG $\wedge^2$ + inhibin-A $\wedge^2$	0.7
	PAPP-A $\wedge^2$ + total hCG $\wedge^2$ + uE3 $\wedge^2$	0.5
	PAPP-A $\wedge^2$ + inhibin-A $\wedge^2$ + uE3 $\wedge^2$	0.3
Four measures + NT	PAPP-A $\wedge^2$ + total hCG $\wedge^2$ + NT	0.6
	PAPP-A $\wedge^2$ + inhibin-A $\wedge^2$ + NT	0.6
	PAPP-A $\wedge^2$ + uE3 $\wedge^2$ + NT	0.3

Fig. 1. Bivariate distribution of log(MoM PAPP-A) in the first trimester (horizontal axis) and second trimester (vertical axis). Circles represent Down's Syndrome outcomes; crosses represent unaffected outcomes. Parameter estimates from Wald *et al.*²

for the Down's Syndrome cases where sample size is relatively small. In addition, the SURUSS parameter estimates are affected by viability bias occurring because of intervention in the second trimester. The effect is likely to optimistically bias assessment of screening performance. Similar biases and over-optimism are likely to affect both the integrated test and the repeated measures tests reported here. Conclusions on the relative merits of these tests must await further prospective and meta-analytic studies, and the performance characteristics presented in this paper should not be quoted out of context.

Table 2. False positive rates for selected tests. In the repeated measures test, it is assumed that first trimester sampling takes place at 10 weeks of gestation. The detection rates and false positive rates relate to the maternal age distribution of England and Wales for the three years 1996 to 1998. It is assumed that MoM values are obtained from scan estimates of gestational age and are corrected for weight. The serum integrated test referred to is the integrated test excluding NT.

Markers included	Detection rate			
	75%	80%	85%	90%
PAPP-A ² , uE3 ² and NT	0.1%	0.2%	0.3%	0.6%
PAPP-A ² and uE3 ²	0.2%	0.3%	0.5%	0.8%
PAPP-A ² , total hCG ² and NT	0.2%	0.3%	0.6%	1.2%
PAPP-A ² and total hCG ²	0.4%	0.7%	1.1%	2.0%
PAPP-A ² , uE3 ² and inhibin ²	0.1%	0.1%	0.3%	0.7%
Quadruple test	2.5%	3.9%	6.2%	10.6%
Combined test	2.3%	3.7%	6.1%	10.8%
Integrated test	0.3%	0.6%	1.2%	2.6%
Serum integrated test	0.8%	1.5%	2.7%	5.3%

However, the results demonstrate that screening tests based on highly correlated repeated measures are likely to have substantial benefits over the integrated test.² More generally, the results provide a cogent demonstration that in the search for new screening tests, we need to cast the net widely to include markers that may not in isolation have good discrimination and that may have high correlations with other markers used in screening.

The estimated properties of the six-marker combination (PAPP-A² + Inhibin-A² + uE3²) raises the interesting possibility of achieving accurate screening results without NT measurements. This may have particular advantages in remote settings such as the Scottish Highlands and Islands where the provision of accurate screening without NT measurements could offer substantial practical advantages. It may therefore be important to confirm the estimated performance of this 'six-marker' test which is more likely to be affected by the sources of bias discussed above.

We have presented results showing that it is possible to improve the performance of the integrated test² by careful selection of markers from both first and second trimesters. However, the tests proposed here share with the integrated test the drawback of requiring that all women, even those with very high or very low estimated Down's Syndrome risks, undergo screening in both the first and second trimesters. A number of authors have argued that such a policy (referred to by Cuckle⁶ as 'non-disclosure' screening) may be unacceptable because of social or medico-legal pressure to reveal 'positive' screening results at first trimester,^{6,7} or even as soon as the NT measurement

has been obtained.⁸ In related work we have examined the possibility of replacing the non-disclosure policy by standard sequential screening, which allows a majority of women carrying normal fetuses to receive results after the first trimester screen is complete. This suggests that up to 75% of pregnancies can be decided at the first trimester with little effect on false positive rates or detection rates.⁹

Finally, it has been argued that abnormal first trimester screening results may preferentially identify fetuses likely to be at risk of early miscarriage, in which case the advantage of integrated or repeated measurement screening over, say, the quadruple second trimester test may be less great than has been claimed. This possibility, together with the acceptability of non-disclosure testing and the relative performance of integrated and repeated measurement screening, should be examined in further prospective studies.

Acknowledgements

The authors would like to thank two anonymous referees and the scientific editor for their constructive comments on an earlier draft of this paper.

References

1. Wald NJ, Watt HC, Hackshaw AK. Integrated screening for Down's syndrome on the basis of tests performed during the first and second trimesters. *N Engl J Med* 1999;**341**(7):461–467.
2. Wald NJ, Rodeck C, Hackshaw AK, et al. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Technol Assess (Winch)* 2003;**7**(11):1–77.
3. Wright DE, Bray I. Estimating birth prevalence of Down's syndrome. *J Epidemiol Biostat* 2000;**5**(2):89–97.
4. Office for National Statistics. *Birth Statistics, Series FM1, Numbers* 25–27, 1998 [1996, 1997, 1998].
5. Aitchison J, Dunsmore IR. *Statistical Prediction Analysis*. Cambridge: Cambridge University Press, 1975.
6. Cuckle HS. Growing complexity in the choice of Down's syndrome screening policy. *Ultrasound Obstet Gynecol* 2002;**19**(4):323–326.
7. Copel JA, Bahado-Singh RO. Prenatal screening for Down's syndrome—a search for the family's values. *N Engl J Med* 1999;**341**(7):521–522.
8. Herman A, Dreazen E, Tovbin J, Weinraub Z, Bukovsky Y, Maymon R. Comparison between disclosure and non-disclosure approaches for trisomy 21 screening tests. *Hum Reprod (Oxf)* 2002;**17**(5):1358–1362.
9. Wright DE, Bradbury I, Benn P, Cuckle H, Ritchie K. Contingent screening for Down's syndrome is an efficient alternative to non-disclosure sequential screening. *Prenatal Diagnosis*. In press.

Accepted 16 June 2004