PRENATAL DIAGNOSIS
Prenat Diagn 2010; **30**: 459–466.
Published online 6 April 2010 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/pd.2502

Second and first trimester estimation of risk for Down syndrome: implementation and performance in the SAFER study

Andrew R. MacRae[1]*†, Bernie N. Chodirker[1], Gregory A. Davies[2], Glenn E. Palomaki[3], George J. Knight[3], Jane Minett[4], Peter A. Kavsak[5], Ants Toi[6], David Chitayat[7] and Paul G. Van Caeseele[8]

[1]*Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada*
[2]*Obstetrics and Gynecology, Queen's University, Kingston, ON, Canada*
[3]*Pathology and Laboratory Medicine, Women & Infants Hospital, Providence, RI, USA*
[4]*Toronto, ON, Canada*
[5]*Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada*
[6]*Diagnostic Imaging, Mount Sinai Hospital, University of Toronto, ON, Canada*
[7]*The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, ON, Canada*
[8]*Medical Microbiology, University of Manitoba, Winnipeg, MB, Canada*

Objectives Document patient choices and screening performance (false positive and detection rates) when three improved Down syndrome screening protocols were introduced coincidentally.

Method Second-trimester 'triple marker' screening was expanded by adding second-trimester dimeric inhibin-A (four-marker), with or without first-trimester pregnancy-associated plasma protein-A (five-marker). Nuchal translucency (NT) measurements were included when available from accredited sonographers (six-marker). For assigning risk, two sets of marker distribution parameters were evaluated.

Results Over 3.5 years, 8571 women enrolled (median age 30.6 years). Uptake of the four-, five- and six-marker protocols was 18%, 46% and 36%, respectively. Of those selecting an integrated test (five or six markers), 9.7% did not provide the second trimester serum sample. False positive rates decreased with added markers (5.2%, 5.1% and 2.5%, respectively) and varied between the two parameter sets, while detection remained high. Overall, 21 of 23 cases were detected (91%, 95% CI 73–98%) at a 4.2% false positive rate (95% CI 3.3–5.1%).

Conclusions Integrated screening protocols were chosen 4.6 times more often than four-marker screening (82% vs. 18% uptake). Overall detection was higher and false positives lower, consistent with recent guidelines. Important performance factors include gestational dating method, risk cut-off, and the parameter set used to assign risk. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS: Down syndrome; prenatal screening; serum markers; nuchal translucency; risk algorithms

INTRODUCTION

Currently, the most efficient prenatal screen for Down syndrome is the full integrated test, a combination of first and second trimester ultrasound and maternal serum markers. Intervention trials showing the effectiveness of integrated screening have been more limited. The original proposal for integrated screening was a modeling exercise based on reliable distributions of existing markers, including AFP, uE3, hCG and DIA in the early second trimester, and PAPP-A and nuchal translucency (NT) in the late first trimester (Wald *et al*., 1999). The SURUS Study (Wald *et al*., 2003 2004, 2006) explored NT measurements in a large cohort, but used a case control format for the stored sera. The aims were to refine parameters and expected performance of the integrated test in relation to other possible combinations in an unbiased manner. The FaSTER Trial (Malone *et al*., 2005) compared first versus second trimester screening algorithms within the same pregnancies, and retrospectively combined markers to assess the integrated test. Neither SURUSS nor FaSTER actually reported integrated risks to the enrolled women, and therefore, could not directly observe the decision-making process.

The first report of an intervention trial of the full integrated test (NT plus five serum markers) was from a centre participating in SURUSS (Pandya *et al*., 2006) and this was subsequently expanded to include a second centre (Wald *et al*., 2009). Other forms of the integrated test have been examined in two other intervention trials: a state-wide program in Maine (Knight *et al*., 2005) introduced the serum integrated test where four-marker screening was established but NT measurements were not widely available, and an urban Toronto program (Okun *et al*., 2008) where access to NT measurement

*Correspondence to: Dr. Andrew R. MacRae, Cadham Provincial Laboratory, 750 William Avenue, Winnipeg, MB, Canada.
E-mail: macrae@cc.umanitoba.ca
† Preliminary data from this study was presented as an unpublished poster at the Annual Clinical Meeting of the Society of Obstetricians and Gynaecologists of Canada, 2008.

Received: 3 November 2009
Revised: 15 January 2010
Accepted: 3 February 2010
Published online: 6 April 2010

was not a limitation, used a version of the integrated test without inhibin-A.

These studies examined single screening algorithms, with access to NT measurements being either essentially uniformly present or absent. In practice, multiple screening modalities are likely to co-exist, and access to NT measurements will vary, particularly when screening services are offered in a wider geography. In addition, screening based on second trimester markers alone continues to be an option for patients presenting too late for first trimester tests or for those declining multiple serum samplings. The second and first trimester estimation of risk (SAFER) study examined patient selection patterns and second trimester screening performance when three Down syndrome screening tests, based on markers from one or both trimesters, were simultaneously implemented in a broad geographical region with evolving access to specialized ultrasound services for NT measurements. Regional and temporal changes in patient test choices could be tracked, along with the performance of the various second trimester screening tests and the two parameter sets used for calculating risk.

METHODS

Study populations and markers

This intervention trial was implemented with approvals from local Research Ethics Boards. First-trimester maternal serum PAPP-A (DSL polyclonal, Webster, TX) and second-trimester inhibin-A (DSL) testing was made uniformly available in the province of Manitoba and in south-eastern Ontario, Canada. In both of these regions, only second-trimester triple marker screening (AFP, uE3 and hCG; Beckman Coulter, Chaska MN) was available at the time the study commenced. Ultrasound measurement of first-trimester NT was incorporated when sonographers certified by the Fetal Medicine Foundation of Canada became available (www.mfmedicine.com). Women presenting in the first trimester were offered either the 'six-marker' integrated screen (all markers combined to yield a second-trimester risk), or the 'five-marker' integrated (no NT) screen depending on NT availability, or the option of waiting until the second trimester. If the woman presented in the second trimester, the 'four-marker' screen (no NT, no PAPP-A) was offered. First-trimester screening was not included as a study option, nor was it performed in the participating regions because the follow-up chorionic villus sampling (CVS) was not available in all study regions. Healthcare provider education was available through a series of educational events in the communities served, after which supplies of patient information and consent documents were distributed. Provider and patient education material concerning the expected screening performance was based on recently modeled retrospective data with added markers (see Assigning Down syndrome risk). The stated goal of the study was to reduce false positive rates while maintaining or increasing detection compared to the existing triple marker screen through the addition of new markers. A description of the logistics of the three study options with the requirement for two scheduled specimens for the integrated screens was a major component of the consent. In south-eastern Ontario, the study was offered in a region of small urban centres and surrounding rural populations, with Queen's University in Kingston as a major educational resource. The majority of the NT measurements in this region were performed at Kingston General Hospital, with other small town sources consistently contributing. Manitoba patients were recruited from the entire province, although NT measurements were only available (Health Sciences Centre, Winnipeg) through an established province-wide referral program for women of advanced maternal age.

Patients were considered to be enrolled in the study if they submitted evidence of consent and a serum specimen before 21 weeks' gestation. Multiple gestation pregnancies were excluded from the study, regardless of when the identification occurred. Uptakes of the three study screening protocols were calculated on the basis of patient intent at enrollment, regardless of whether the selected screen was completed. Demographic and pregnancy-related information collected for risk calculation included racial origin, diabetic status, maternal weight and smoking. When gestational ages were based on CRL or BPD measurements, they were assigned according to published equations (Daya, 1993; Hadlock *et al*., 1982). Otherwise, the gestational age was nearly always based on last menstrual period.

Study protocols and compliance

Receipt of a serum specimen from both the first and second trimesters (between $8+0$ and $13+6$ weeks, and between $15+0$ and $20+6$ weeks, respectively), triggered a five-marker (serum integrated) screen, unless an NT measurement was also provided, in which case, the six-marker (full integrated) screen was performed. No Down syndrome risk was provided if essential data were not available (e.g. NT alone, or no second trimester specimen). Receipt of only a second trimester specimen triggered a four-marker (quadruple) screen. When gestational age revision after initial screening caused a specimen to be too early or too late for the initial protocol, other study protocols were substituted, with instructions for repeat collections as required. The populations in the initial and final screens for a given protocol were therefore slightly different. Integrated screening compliance was monitored and reminders for missing second trimester specimens were issued after 17 completed weeks. Integrated screens were deemed incomplete when no valid second trimester serum specimen was received by 20 weeks, 6 days. In assessments of compliance, a screening test was considered to have been waived if amniocentesis was performed prior to the day the first reported risk was calculated.

Assigning Down syndrome risk

All markers were expressed in multiples of the median (MoM) and adjustments were made for maternal weight,

DOI: 10.1002/pd

diabetes (AFP/0.8) and Black racial origin (AFP/1.10). No median adjustments were made for smoking. Software for the study was developed by Benetech Medical Systems, Toronto. The Down syndrome risk at term was computed using the age-associated prior risk (Cuckle *et al*., 1987), multiplied by a likelihood ratio computed from multivariate overlapping Gaussian distributions obtained from the literature. During separate phases of the study, two parameter sets were evaluated for assigning risk. During Phase I ($n = 4681$), the parameters used for the four-marker screen were from one UK dataset (Wald *et al*., 1994, 1996, 2000), and the PAPP-A and NT parameters were from the same group in the UK (Wald and Hackshaw, 1997). Subsequently in Phase II ($n = 3890$), designed to allow comparison of performance with a different parameter set, the AFP, uE3 and hCG parameters (Knight *et al*., 1998), inhibin-A (Haddow *et al*., 1998a) and PAPP-A (Haddow *et al*., 1998b) were from a USA group, while the NT parameters were from a different group in the UK (Nicolaides *et al*., 1998). Down syndrome term risk cut-offs were 1 : 260 for all three screening tests in Phase I, and in Phase II slightly higher term risk cut-offs were used, 1 : 250 for the four-marker screen and 1 : 200 for both the five- and six-marker screens. The higher risk cut-offs implemented in Phase II reflected evolving practice in Canada and were selected after modeling the maternal age distribution with the comparator parameter set to reduce the false positive rate with minimal decrease in detection. There were no other differences in design between Phases I and II.

In both phases, in accordance with the published parameter sets implemented, there was no correlation between NT and any biochemical marker, and there was no change in the affected population mean by week for PAPP-A or NT. This was consistent with early-integrated screening practice in North America in 2003 when the study began.

Follow-up and outcome ascertainment

Follow-up counselling and healthcare options (e.g. confirmation of clinical information, ultrasound examination and amniocentesis) were available through primary providers and at regional genetic centres. Updates to clinical information such as gestational age were performed according to standard protocol (e.g. a minimum 10-day difference was required for amendment). Screening for disorders other than Down syndrome (open neural tube defects, trisomy 18 and Smith-Lemli-Optiz syndrome—SLOS) proceeded in the established manner (Palomaki *et al*., 1995, 2002), and patients who screened positive for these disorders continued in the study. Compliance, follow-up choices, outcomes and prenatal and post-delivery diagnostic test results were collected through four overlapping processes: (1) from healthcare providers when patients were late or did not submit required specimens, (2) from the three referral centres for karyotype analyses that served the study, (3) directly from approximately 670 healthcare providers serving study patients via mailed/faxed pregnancy-specific outcome gathering forms, and (4) from logs of serum specimens collected coincidentally with each amniocentesis performed. Down syndrome was confirmed by cytogenetic analysis in 23 cases and by autopsy in 2 cases. The number of expected cases at term based on the maternal age distribution in the study was calculated according to published models (Morris *et al*., 2005; Morris and Wald, 2007). Ascertainment at mid-trimester was calculated based on a 23% spontaneous loss rate for the affected cases between second trimester and term (Bray and Wright, 1998; Morris *et al*., 1999; Morris and Wald, 2007).

Screening test performance

All enrolled pregnancies with calculated Down syndrome risks were used in the initial and final false positive and detection rate calculations. To compare the false positive rates of the two sets of marker distribution parameters and risk cut-off levels (Phases I and II), risks were retrospectively calculated based on both parameter sets for all available markers in each pregnancy. Prospectively observed Down syndrome detection for the overall study was directly computed and then adjusted for bias of ascertainment. Down syndrome detection and false positive rates are reported with exact 95% confidence intervals (TrueEpistat, Richardson, TX).

RESULTS

Between January 2003 and June 2006, a total of 8571 singleton pregnancies were enrolled in the SAFER study. Table 1 shows demographic and pregnancy-related information for these women, stratified by the initial screening method completed. Important differences exist, mainly between the six-marker and the other two protocols. Women completing the six-marker test were older (27% age 35 or older compared to 15% in the other groups, $\chi^2 = 177$, $p < 0.001$) and all received a gestational age estimate based on CRL as part of the measurement of NT (over half of women in the other groups were dated via last menstrual period, $\chi^2 = 6000$, $p < 0.001$). Women receiving the six-marker test were also more likely to be Caucasian ($\chi^2 = 77$, $p < 0.001$), and had a lower rate of smoking ($\chi^2 = 25$, $p < 0.001$) than those receiving the four- or five-marker tests.

Table 1 also shows that the overall uptake rates for the three screening protocols were 18%, 46% and 36% for the four-, five- and six-marker protocols. These uptake rates are stratified by region and quartiles of recruitment in Table 2. Within Manitoba, 74% of recruits chose integrated screening, and the choices did not change appreciably over time, once NT became available as an addition to the existing program for women of advanced maternal age. NT was thereafter uniformly available in Manitoba through this referral practice, and uptake of six-marker was the same (11%) in rural and large urban

DOI: 10.1002/pd

Table 1—Demographic and pregnancy-related information for the women enrolled in the SAFER study, stratified by the screening protocol initially chosen

Parameter	Four-marker	Five-marker	Six-marker	All
Number	1580 (18%)	3897 (46%)	3094 (36%)	8571
Age at delivery[a]	29.3 (24.8, 33.4)	30.3 (26.5, 33.7)	31.8 (27.7, 35.6)	30.6 (26.6, 34.4)
Age ≥35 (%)	14.6	15.1	26.8	19.2
Primagravida (%)	37.6	33.3	33.8	34.4
Weight (kg)[a]	68.0 (59.0, 80.7)	68.0 (59.4, 79.8)	67.6 (59.9, 79.4)	68.0 (59.4, 79.8)
Gestational age based on				
CRL (%)	25	32	100	55
BPD (%)	19	7	0	7
Dates (%)	56	61	0	38
Racial origin self-declared as				
Caucasian (%)	80	86	91	87
Black (%)	1	1	1	1
Asian (%)	9	7	6	7
First Nation (%)	9	5	1	4
Other (%)	1	1	1	1
IDDM (%)	0.8	0.9	1.3	1.0
Smoking (%)	17.3	13.4	10.7	12.2

CRL, crown rump length; BPD, biparietal diameter; IDDM, insulin dependent diabetes mellitus;
[a] Median (25th, 75th centiles).

Table 2—Choice of screening protocol by region and quartiles of enrolment

	Number of women (%)		
Quartile	Manitoba	SE Ontario	All
Four-marker			
First	323 (38)	216 (17)	539 (25)
Second	203 (25)	178 (14)	381 (18)
Third	212 (25)	155 (12)	367 (17)
Fourth	144 (17)	149 (12)	293 (14)
All	882 (26)	698 (14)	1580 (18)
Five-marker			
First	521 (61)	335 (26)	856 (40)
Second	559 (64)	432 (33)	991 (46)
Third	545 (64)	465 (36)	1010 (47)
Fourth	615 (72)	425 (33)	1040 (48)
All	2240 (66)	1657 (32)	3897 (46)
Six-marker			
First	8 (1)	740 (57)	748 (35)
Second	90 (11)	681 (53)	771 (36)
Third	95 (11)	671 (52)	766 (36)
Fourth	92 (11)	717 (55)	809 (38)
All	285 (8)	2809 (54)	3094 (36)
	3407	5164	8571

settings. In south-eastern Ontario, the rate of six-marker testing remained relatively constant at 54%, while the preference in other recruits migrated over time, away from four-marker and toward five-marker (17% and 26% in the first quartile, to 12% and 33% in the fourth quartile).

First-trimester serum specimens were accepted as early as 8 weeks, 0 days, but informational materials listed 9 weeks, 0 days in order to reduce the frequency of specimens drawn too early, as observed by others (Knight *et al.*, 2005). The six-marker protocol did not require first-trimester blood sampling and NT measurement at the same time. Although earlier serum collection was encouraged, 88% of the women had their blood drawn on the same date as their NT measurement. Another 9% had blood drawn prior to the NT being measured (95% of these within 3 weeks) and 3% had blood drawn afterwards.

Some women did not complete the screening protocol that they initially chose. Table 3 provides information about the reasons for non-completion as well as the final protocol used for interpretation. A small proportion of women (0.9%) chose diagnostic testing prior to receiving the results of their screening test. An additional 9.7% of those choosing the integrated protocols did not provide the required second-trimester specimen. The reasons for this were collected during and following the process of issuing reminders to healthcare providers when the second-trimester specimen had not been received by 18 weeks (30% of all integrated screens required this reminder). The three most common causes for incomplete integrated screens were spontaneous fetal loss (2.3%), withdrawal from the study (2.1%) and unspecified reasons (1.4%). Another 1% either required or were candidates for a different screen from their original choice, often due to discrepancies in the initial gestational age assignment based on LMP. Cumulatively, 11.3% of patients choosing integrated protocols did not receive their 'entry' screen, compared to 1.6% of those choosing the four-marker protocol.

As specified in the study plan, the overall false positive rate was monitored during the initial study enrolment and compared to the rate that would have been found had those pregnancies been screened using the triple test. After 442 patients, there were 20 screen positive pregnancies (overall observed screen positive rate of 4.5%; 95% CI 2.9–6.9%). This was significantly lower than the expected 41 pregnancies had these women received the triple test (screen positive rate 9.3%; 95% CI 6.9–12.4%).

DOI: 10.1002/pd

Table 3—Uptake and compliance with the Down syndrome screening protocols initially chosen, along with the protocol used for final interpretation

	Down syndrome screening protocol			
	Four-marker	Five-marker	Six-marker	Any
Subjects initially choosing protocol	1580	3897	3094	8571
Amniocentesis before interpretation	5	46	23	74
No second trimester specimen submitted	NA[a]	415 (10.5%)	263 (8.5%)	678 (9.7%)
Suiting or requiring a different protocol	21	43	0	64
Did not receive the 'entry' screen	26 (1.6%)	504 (12.9%)	286 (9.2%)	816 (9.5%)
Transfers to this protocol	41	15	8	
Final screen by this protocol	1595	3408	2816	7755

[a] NA, not applicable.

Table 4—Initial and revised false positive rates, stratified by Down syndrome screening protocol and study phase

			Initial false positive rate				Revised false positive rate			
Protocol	Study phase[a]	Cut-off (1 : n)	UA	Positive	FPR (%)	All[b] (%)	UA	Positive	FPR (%)	All[b] (%)
Four-marker	Phase I	260	963	89	9.2		955	61	6.4	
	Phase II	250	634	45	7.1	8.4	628	21	3.3	5.2
Five-marker	Phase I	260	1756	140	8.0		1752	107	6.1	
	Phase II	200	1691	112	6.6	7.3	1682	69	4.1	5.1
Six-marker	Phase I	260	1715	53	3.1		1720	53	3.1	
	Phase II	200	1111	18	1.6	2.5	1114	18	1.6	2.5
All[c]			7870	457		5.8 95% CI 5.2–6.3	7851	329		4.2 95% CI 3.7–4.6

UA, unaffected pregnancy; FPR, false positive rate;
[a] Between Phases I and II, the marker distribution parameter sets and the Down syndrome risk cut-off levels were changed;
[b] Both phases combined;
[c] Numbers of initial and revised screens do not match due to gestational age changes causing removal or changed protocol.

In the overall study ($n = 8571$), the initial false positive rate (FPR) was 5.8% of screened pregnancies and after ultrasound reclassification of gestational age, the rate dropped to 4.2%. Table 4 stratifies these rates by screening protocol and by study phase. For all three protocols, the risk cut-off level was increased in Phase II, and the distribution parameters used to assign risk were changed as described in the section on Methods. Several important findings emerge. In both study phases, the false positive rates are significantly lower after ultrasound reclassification for the four- and five-marker protocols (where more than half of the pregnancies were dated by LMP), but there was no change for the six-marker protocol (all pregnancies dated by CRL). Within each phase, the revised false positive rates for the four- and five-marker protocols were similar, but the revised false positive rate for the six-marker protocol was considerably lower.

The changes implemented for Phase II had varying effects on the false positive rates for the three protocols. For the four-marker test, Phase II is associated with a large reduction in the false positive rate (revised rates drop from 6.4% in Phase I to 3.3% in Phase II), even though the risk cut-off did not change appreciably (1 : 260 to 1 : 250). The change in parameter sets is likely responsible. For the five-marker test, the rates also drop (revised positive rates 6.1 to 4.4%), but it is less clear whether this was caused by the higher risk cut-off (1 : 260 to 1 : 200) or the parameter set changes. We reanalyzed the Phase I data using a 1 : 200 cut-off level and found a revised rate of 4.6%, indicating that most of the reduction was due to the change in the risk cut-off level. A similar reanalysis was performed for the six-marker test where the revised rates had dropped from 3.1% in Phase I to 1.6% in Phase II. Retrospectively applying the 1 : 200 cut-off to the Phase I data returned a false positive rate of 2.4%, indicating that the reduction in false positive rate is only partially explained by the change in cut-off, and the change in the parameter set is also a factor.

There were 25 cases of Down syndrome ascertained in the SAFER population (Table 5), and all were detected prior to 30 weeks' gestation. Two patients with affected pregnancies chose to have CVS on the basis of abnormal first trimester ultrasound examinations (NT measurements of 7.0 and 4.8 mm; 4.98 and 3.36 MoM, respectively). No second trimester specimens were submitted, although the pregnancies were continuing. The remaining 23 cases of Down syndrome had a risk assigned by one of the three protocols; 21 were screen positive and 2 were screen negative. One screen negative case (Table 5 #15; five-marker) had increased risk of SLOS and miscarried at 19 weeks. The other (#25, four-marker) had a 46,XX,der(21;21)(q10;q10) translocation

DOI: 10.1002/pd

Table 5—Down syndrome pregnancies identified in the SAFER study

Case #	Age at delivery	Multiple of the median (MoM) NT	PAPP-A	AFP	uE3	hCG	DIA	Karyotype	Term DS risk (1 : *n*)	GA at Dx (weeks, days)
Incomplete six-marker tests										
1	33.0	4.98	0.68	NA	NA	NA	NA	47,XX,+21	NA	12, 4
2	46.9	3.36	0.29	NA	NA	NA	NA	47,XX,+21	NA	12, 3
Six-marker tests										
3	29.7	5.05	0.59	0.88	0.90	1.83	1.56	47,XX,+21	<8 (I)[a]	16, 4
4	33.9	2.63	1.23	0.33	0.91	3.16	1.47	47,XY,+21	<8 (I)	38, 4
5	35.7	3.59	1.38	1.14	0.65	4.17	2.56	46,XX,der(21;21)(q10;q10)	<8 (I)	15, 2
6	40.4	1.33	1.06	0.66	0.74	2.50	1.79	47,XY,+21	15 (I)	20, 6
7	36.3	1.82	0.47	1.65	0.80	2.22	4.91	NA[b]	33 (II)	22, 2
Five-marker tests										
8	27.9	NA	0.36	1.72	0.75	3.41	2.84	47,XX,+21	<8 (I)	22, 5
9	31.4	NA	0.09	0.35	0.34	3.42	1.93	47,XY,+21	<8 (I)	16, 6
10	41.4	NA	1.44	0.85	0.67	2.71	2.79	47,XX,+21	11 (I)	40, 6
11	36.5	NA	0.74	1.29	0.56	2.18	3.63	47,XY,+21	44 (I)	16, 1
12	35.0	NA	0.46	0.68	0.45	1.90	0.92	47,XX,+21	48 (II)	19, 3
13	32.3	NA	0.85	0.60	1.38	5.95	4.49	47,XX,+21	54 (II)	38, 5
14	37.8	NA	0.70	0.81	0.69	1.81	1.33	47,XY,+21	118 (I)	38, 2
15	36.0	NA	0.59	1.85	0.17	0.92	0.85	47,XX,+21	106,000[c] (I)	19, 6
Four-marker tests										
16	44.0	NA	NA	1.25	0.74	9.22	10.27	NA[d]	<8 (I)	39, 0
17	39.8	NA	NA	0.96	0.73	3.80	2.10	47,XY,+21	<8 (I)	16, 1
18	31.6	NA	NA	0.56	0.22	4.94	1.03	47,XX,+21	<8 (I)	19, 2
19	45.8	NA	NA	0.58	0.92	3.27	4.71	47,XY,+21	<8 (I)	22, 4
20	36.9	NA	NA	0.73	0.96	2.83	2.25	47,XY,+21	25 (I)	17, 3
21	28.8	NA	NA	0.57	0.69	1.81	2.74	46,XY,der(21;21)(q10;10)	37 (I)	17, 3
22	32.7	NA	NA	0.56	0.79	1.84	1.81	47,XY,+21	66 (I)	16, 4
23	35.3	NA	NA	0.62	0.64	1.60	1.59	47,XY,+21	68 (II)	19, 5
24	37.0	NA	NA	0.49	0.79	0.93	1.90	47,XY,+21	81 (II)	19, 5
25	29.8	NA	NA	1.07	0.88	1.70	1.46	46,XX,der(21;21)(q10;q10)	1770[e] (II)	29, 4

NT, nuchal translucency; PAPP-A, pregnancy-associated plasma protein-A; AFP, alpha-fetoprotein; uE3, unconjugated estriol; hCG, human chorionic gonadotrophin; DIA, dimeric inhibin-A; DS, Down syndrome; GA, gestational age; Dx, diagnosis; NA, not available;
[a] Screened in Phase I (I), or Phase II (II);
[b] Stillbirth at 22 weeks, presumed Down syndrome based on physical features and low acetabular index, post-mortem;
[c] Also positive for SLOS; miscarried at 19 + 6 weeks; the risk of DS using Phase II parameters would have been 1 : 320;
[d] Declined amniocentesis; IUD at 39 weeks; autopsy anomalies support Down syndrome; blood karyotype failed;
[e] The risk of DS using Phase I parameters would have been 1 : 1020.

and was detected at 29 weeks because of an abnormal ultrasound finding (IUGR). Of the 16 cases of Down syndrome with a karyotype determined prior to 23 weeks, 12 (75%, 95% CI 48–93%) chose to terminate the pregnancy. Based on the maternal age distribution of the enrolled women, 24.5 mid-trimester cases of Down syndrome would be expected (Morris *et al.*, 2005; Morris and Wald, 2007), assuming a natural fetal loss rate of 23% between second trimester and term in affected pregnancies (Bray and Wright, 1998; Morris *et al.*, 1999). Of the six cases ascertained in the late third trimester and at term, five were screen positive (equivalent to 6.5 of 7.8 mid-trimester cases). In the 17 cases ascertained at mid-trimester, 16 were screen positive. Thus, after correction for trimester of ascertainment, the Down syndrome detection rate observed with the combination of screening protocols selected by patients was 22.5 of 24.8 or 91% (approximate 95% CI: 73–98%). The odds of being affected given a positive result (OAPR) were approximately 1 : 15 (22.5 : 329).

DISCUSSION

The SAFER study demonstrates that a high detection rate and acceptably low false positive rate can be achieved by offering an array of screening protocols across a broad rural and urban geography. Guidelines for prenatal screening in America, Canada and UK recently adopted goals for reducing false positive rates below 5% while maintaining or increasing detection of Down syndrome (ACOG, 2007; Summers *et al.*, 2007; NICE, 2008). The overall performance in the study met these goals by offering up to three protocols to patients and with risk cut-off values that evolved during the study. In the study regions, triple-marker screening was the common, government-funded standard of care when the SAFER study commenced. For that reason, the four-marker screen was the easiest option for healthcare providers, requiring no change in practice. The four-marker option was also the only study protocol available to patients presenting after 13 completed weeks. Despite this, uptake of the four-marker screen was the lowest

DOI: 10.1002/pd

of the three made available by the study. Selection of integrated screening protocols with their requisite two serum specimens at differing gestational ages predominated throughout the study. If combined first trimester screening and CVS had also been available, presumably, the proportion of integrated screens would have been less. Because of the wide geography and varying practice settings included in our study, region-specific differences were clear. In one region, for example, NT measurements were mainly reserved for women of advanced maternal age, while in the other region they were more routinely available. Where NT measurements were less readily available, five-marker integrated screening was chosen by two-thirds of the enrolees. Over the entire study population, five-marker uptake was 46%, which is considerably higher than the 5% uptake reported in a localized urban setting (Okun *et al*., 2008), perhaps reflecting the more varied and general practice settings included in our study.

An important fraction (9.7%) of women initially selecting an integrated screening protocol did not complete any screening protocol, consistent with rates from similar studies (Knight *et al*., 2005; Pandya *et al*., 2006; Okun *et al*., 2008). Further investigation is warranted to see if women are fully informed of the requirements to complete integrated screening. Offering six-marker screening in a stepwise or sequential manner may provide a higher proportion of women with acceptably reliable risks. In some instances such 'opting out' seems warranted. For example, among the 14 pregnancies with an NT measurement greater than 3.5 mm (0.45% of all NT measurements), there were four cases of Down syndrome, two of trisomy 18, one Turner syndrome, five normal karyotypes and one unknown. Unfortunately, such stepwise screening is not feasible without the NT marker, and it was the five-marker serum-only protocol that had the highest rate of non-completion.

Estimating confident protocol-specific Down syndrome detection rates was not a goal of the study. Because this was an intervention trial with ascertainment of cases at different times in gestation, it would be difficult to provide reliable rates, even if the study were larger. It is noteworthy, however, that all cases of Down syndrome were identified before birth by either the three prenatal screening tests or the associated standard obstetrical care including later ultrasound examinations.

As many previous studies have shown, our data confirm that ultrasound correction of gestational age will result in much lower false positive rates (Table 4), with the expectation that the Down syndrome detection rate will be similar, or higher. Over half of the enrolled women choosing four- and five-marker tests had their pregnancy dated by last menstrual period. Routine ultrasound-based dating prior to serum test interpretation would improve performance and should be strongly encouraged. Through interim analyses during the study, we determined that sufficient data existed to warrant increasing the risk cut-off levels to significantly reduce the false positive rate with the expectation that the detection rate would be only slightly reduced. The higher risk cut-offs were modeled with the comparator parameter set, and implemented in Phase II to assign patient-specific Down syndrome risks. One unexpected consequence was the important difference in the false positive rates for the four-marker screen, depending on the parameter set chosen. Using the limited number of Down syndrome cases in our study, this reduction in false positive rate did not appear to reduce detection. This finding deserves further study to determine which parameter(s) might be causing the effect.

The screening performance summarized in this report could be improved further by two additional refinements: adjusting PAPP-A and DIA MoM levels for maternal smoking status, and varying the first trimester Down syndrome population parameters by gestational week (Bindra *et al*., 2002; Wald *et al*., 2004). Although this latter concept was evolving in first trimester screening publications at the time of the study, it was not part of either of the distribution parameter sets used in our study, nor was it yet in use in any second trimester integrated screening programs.

Initial and revised false positive rates are dependent on the combination of markers, risk cut-off levels, the distribution of maternal ages, method of dating the pregnancy, rules for reclassifying 'bad dates', the parameter set chosen to assign risks, the gestational age range for first-trimester results, and many other factors; however, comparisons that are performed within our study population will be reasonably robust. If the study population had received triple marker screening in both phases, the overall revised screen positive rate would have been 8.9% (372 more potential amniocentesis procedures) with four false negatives instead of two. In the published literature, the most relevant comparison would be our five-marker results, stratified by dating method, with those from a similar state-wide intervention trial (Knight *et al*., 2005). Using similar cut-off levels and parameter sets, they reported initial and revised false positive rates of 4.1% and 3.5% for LMP dated pregnancies (our corresponding Phase II rates were 6.6% and 4.1%). The somewhat higher rates in our study may be due to our older population (average age 31.3 vs. 27.8 years; 15.3% vs. 11.3% age 35 or older).

Down syndrome screening programs offering the four-, five- and six-marker algorithms in a general pregnancy population can meet the North American and UK guidelines for high detection at correspondingly low false positive rates.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Canadian Institutes of Health Research (MOP62884) and by an unrestricted grant from Beckman Coulter of a portion of the analytical reagents. We thank Gail Patterson and Kym Peixoto for their diligence in the screening service, and we acknowledge the analytical expertise of Louis M. Neveux, Sharon Janushewski, Jan Landin, Flora Cantlon and Evelyn Moreau.

REFERENCES

ACOG Committee on Practice Bulletins. 2007. ACOG Practice Bulletin No. 77: screening for fetal chromosomal abnormalities. *Obstet Gynecol* **109**: 217–227.

Bindra R, Heath V, Liao A, Spencer K, Nicolaides KH. 2002. One-stop clinic for assessment of risk for trisomy 21 at 11–14 weeks: a prospective study of 15 030 pregnancies. *Ultrasound Obstet Gynecol* **20**: 219–225.

Bray IC, Wright DE. 1998. Estimating the spontaneous loss of Down syndrome foetuses between the times of chorionic villus sampling, amniocentesis and livebirth. *Prenat Diagn* **18**: 1045–1054.

Cuckle HS, Wald NJ, Thompson SG. 1987. Estimating a woman's risk of having a pregnancy associated with Down's syndrome using her age and serum alpha-fetoprotein level. *Br J Obstet Gynaecol* **94**: 387–402.

Daya S. 1993. Accuracy of gestational age estimation by means of fetal crown-rump length measurement. *Am J Obstet Gynecol* **168**: 903–908.

Haddow JE, Palomaki GE, Knight GJ, Foster DL, Neveux LM. 1998a. Second trimester screening for Down's syndrome using maternal serum dimeric inhibin A. *J Med Screen* **5**: 115–119.

Haddow JE, Palomaki GE, Knight GJ, Williams J, Miller WA, Johnson A. 1998b. Screening of maternal serum for fetal Down's Syndrome in the first trimester. *New Engl J Med* **338**: 955–961.

Hadlock FP, Deter RL, Harrist RB, Park SK. 1982. Fetal biparietal diameter: a critical re-evaluation of the relation to menstrual age by means of real-time ultrasound. *J Ultrasound Med* **1**: 97–104.

Knight GJ, Palomaki GE, Neveux LM, Fodor KK, Haddow JE. 1998. hCG and the free β-subunit as screening tests for Down Syndrome. *Prenat Diagn* **18**: 235–245.

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

Malone FD, Canick JA, Ball RH, *et al*., for the First- and Second-Trimester Evaluation of Risk (FASTER) Research Consortium. 2005. First-trimester or second-trimester screening, or both, for Down's syndrome. *New Engl J Med* **353**: 2001–2011.

Morris JK, Wald NJ. 2007. Estimating the risk of Down's syndrome in antenatal screening and the gestation at which this risk applies. *J Med Screen* **14**: 5–7.

Morris J, Mutton D, Alberman E. 2005. Correction to maternal age-specific live birth prevalence of Down's syndrome. *J Med Screen* **12**: 202.

Morris JK, Wald NJ, Watt HC. 1999. Fetal loss in Down syndrome pregnancies. *Prenat Diagn* **19**: 142–145.

NICE 2008. *Antenatal Care—Routine Care for the Healthy Pregnant Woman*. RCOG Press: London.

Nicolaides KH, Snijders RJM, Cuckle HS. 1998. Correct estimation of parameters for ultrasound nuchal translucency screening. *Prenat Diagn* **18**: 519–521.

Okun N, Summers AM, Hoffman B, *et al*. 2008. Prospective experience with integrated prenatal screening and first trimester combined screening for trisomy 21 in a large Canadian urban center. *Prenat Diagn* **28**: 987–992.

Palomaki GE, Bradley LA, Knight GJ, Craig WY, Haddow JE. 2002. Assigning risk for Smith-Lemli-Opitz syndrome as part of 2nd trimester screening for Down's syndrome. *J Med Screen* **9**: 43–44.

Palomaki GE, Haddow JE, Knight GJ, *et al*. 1995. Risk-based prenatal screening for trisomy 18 using alpha-fetoprotein, unconjugated oestriol and human chorionic gonadotrophin. *Prenat Diagn* **15**: 713–723.

Pandya PP, Jones P, Huttley W, Rodeck C, Wald N. 2006. Integrated screening for Down syndrome in routine clinical practice. *Ultrasound Obset Gynecol* **28**: 362–363.

Summers AM, Langlois S, Wyatt P, Wilson RD. 2007. Prenatal screening for fetal aneuploidy. *J Obstet Gynaecol Can* **29**: 146–179.

Wald NJ, Densem JW, George L, Mattakrishna S, Knight PG. 1996. Prenatal screening for Down's Syndrome using inhibin A as a serum marker. *Prenat Diagn* **16**: 143–153 [Erratum: *Prenat Diagn* **17**: 285–290].

Wald NJ, Densem JW, Smith D, Klee GG. 1994. Four marker serum screening for Down's Syndrome. *Prenat Diagn* **14**: 707–716.

Wald NJ, Hackshaw AK. 1997. Combining ultrasound and biochemistry in first-trimester screening for Down's syndrome. *Prenat Diagn* **17**: 821–829.

Wald NJ, Hackshaw AK, George LM. 2000. Assay precision of serum α fetoprotein in antenatal screening for neural tube defects and Down's syndrome. *J Med Screen* **7**: 74–77.

Wald NJ, Huttly WJ, Murphy KW, Ali K, Bestwick JP, Rodeck CH. 2009. Antenatal screening for Down's syndrome using the Integrated test at two London hospitals. *J Med Screen* **16**: 7–10.

Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM. 2003. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Tech Assess* **7**: 1–77. (See also J Med Screen **10**: 56–104.).

Wald NJ, Rodeck C, Hackshaw AK, Rudnicka A. 2004. SURUSS in perspective. *BJOG* **111**: 521–531.

Wald NJ, Rodeck C, Hackshaw AK, *et al*. 2006. Errata correction to SURUSS report. *J Med Screen* **13**: 51–52.

Wald NJ, Watt HC, Hackshaw AK. 1999. Integrated screening for Down's syndrome on the basis of tests performed during the first and second trimesters. *New Engl J Med* **341**: 461–467.

DOI: 10.1002/pd