A Comparison Study on First and Second Trimester down'sSyndrome Screening Results in Pregnant Albanian Women

Emil G. Xhuvani[a], Aleksander P. Xhuvani[b], Anila K.Mitre[c]
[a] Department of Biophysics, University of Tirana, Faculty of Natural Sciences, Street Bulevardi "Zog I", Tirana, Albania
[b] Department of Computer Engineering, Polytechnic University of Tirana, Square Mother Teresa, nr 1", Tirana, Albania.
[c]Intermedica Diagnostic Tirana, Street "Rreshit Petrela" Nr. 3, Tirana, Albania

Abstract

Nowadays in Albania pregnant women are screened for fetal Down's syndrome (T21) presence by performing among other test as well the first and second trimester screeningtests.Concentration of PAPP-A, free-β-hCG, AFP, total HCG were measuredwith COBAS6000;μE_3 (oestriol)with ELISA method. NT, CRL and BPD were measured with standardized ultrasound techniques. Calculations T21 risk were computed with ssdlab5. The screening results for T21 were confirmed through pregnancy termination and karyotyping. 514 women with singleton pregnancies underwent combined screening test at 73 days through 97 days of gestation age and 487 women with singleton pregnancies underwent to second-trimester screening (triple test) at 105 days through 140 days of gestation age. In the first-trimester screening we had 3 cases with T21at a 4.7% false positive rate, 3 cases for the triple test at a 9.92% false positive rate and 0% false negative rate for both tests. Sensitivity for combined test was at 100% (CI 30.48%↔100.00%)while for triple test was at 100% (CI 30.48%↔100.00%); specificity for combined test was at 95.30% (CI 93.09%↔96.97%)while for the triple test was at 90.08% (CI 87.07%↔92.60%). Combined test has a better performance than triple test regarding screening for T21.

KEYWORDS: sensitivity, screening test, Down syndrome, T21, specificity.

INTRODUCTION

Nowadays in Albania a big number of pregnant women are screened for fetal Down's syndrome (T21) presence by performing among other test as well the first and second trimester screening tests. The first trimester screening test assess the fetus risk for T21 (and other aneuploidies) by means of the combination ofmaternal serum biomarkers concentration as Pregnancy-Associated Plasma Protein A (PAPP-A), free subunit β of human Chorionic Gonadotropin (free β-hCG) and the nuchal translucency (NT) measurement. The blood drawn from pregnant women for biochemical analysis is performed between 9 and 13 weeks'+6 days of gestation, and on the same day is performed the ultrasound measurement of NT.All these data are computed in a statistical program that asses the risk of the fetus to be or not affected with one of the chromosomal aneuploidies. Detection rate, the ratio between the screen positive cases with true positive cases, for Down syndrome range from 79 to 90% with a 5% false-positive rate*(Spencer, K., et al., 2003).*In the cases of affected fetus with T21 the PAPP-A levels are reduced, and free β-hCG is increased. The sensitivity is increased with the combination of NT

measurements *(Bindra, R. at al., 2002).* This test performed at this gestational age gives better possibilities for later on diagnostics and invasive procedure like chorionic villus sample and amniocentesis. Combined first trimester serum screening for multifetal gestations is less sensitive (70%) than in singleton pregnancies (*Cleary-Goldman J. et al., 2005*).

The second trimester screening test, known also as "triple test", relies on three maternal serum biomarkers like Alfa-fetoprotein (AFP), Human Chorionic Gonadotropin (HCG) and oestriol (uE3).The ultrasound measurement is performed to measure the bi-parietal diameter (BPD) to provide more accurate gestational age. The triple test has a sensitivity of ~65% for T21 for singleton pregnancies andin twin gestations, second trimester serum screening detects only ~50% of affected fetuses *(Neveux LM et al., 1996.* The last menstrual period (LMP) is the common method used to assess maternal gestational age (GA). The ultrasound measurement provides a better assessment for GA through measurement of crown-rump length (CRL) in the first-trimester and of BPD in the second trimester. If the gestational age assessed (with LMP and ultrasound) changes by >10 days then the test results must be reinterpreted (*Palomaki GE et al., 2005*). Women and obstetricians choose for many reasons first or second trimesters screening tests. The aim of this paper is to compare some statistical parameters between these two screening tests. These two screening test dohave differences in performance that can be assessed through false positive rate, false negative rate, sensitivity and specificity. By means of ROC analysis is possible to compare the advantages and disadvantages of both screening test.

MATERIALS AND METHODS

Between 2009 and 2014 in Intermedica clinical laboratory of Tirana a number of 514 women with singleton pregnancies, from different obstetrical centers, underwent combined screening test at 73 days through 97 days of gestation age and 487 women with singleton pregnancies underwent to second-trimester screening (triple test) at 105 days through 140 days of gestation age. The pregnant women population of first and second trimester screening test were partially randomly chosen and partially based on their age which means that the *a priori risk* could be higher than in a large clinical routine unselected population. An application form is compiled to be fulfilled by the obstetrician and the pregnant woman. The used criteria to include in the data base the cases are: gestational age between 9 and 22 weeks (accordingly with the screening test); singleton pregnancy; nonsmoker; maternal age 18 years old or older; pregnant women consent to undergo the blood analysis; all albanian pregnant women; no diabetique and insuline dependent pregnant women; all pregnancies confirmed throught pregnancy termination or karyotiping; biomarkers concentration values and ultrasound measurement were measured within 24 hours; concentration biomarkers values under and above the normal limits are truncated accordingly with the used method.In compliance with local ethics requirements, each woman completed a consent form. All other cases that were missing one the above mentioned criteria are excluded from the data base.

Concentration of PAPP-A, free-β-hCG, AFP, and total HCG were measured with COBAS6000analyzer with sandwich principle with immunoassays kits provided by Roche, andμE_3 concetrations were automatically measured with enzyme-linked immunosorbentassay (ELISA) by means of Tecan SUNRISE by measuring the optical density and 450nm. NT, CRL and BPD were measured with standardized ultrasound

techniques. All results were expressed as Multiple of Medians (MoM).Calculations T21 risk were computed with ssdlab5. The statistical parameters like false positive and negative, sensitivity and specificity are computed by means of "null" hypothesis and ROC (receiver operator characteristic).The comparison of sensitivity between two screening test at a fix specificity or sensitivity level has been carried out also with ROC analysis *(DeLong et al., 1988)*. In the absence of the false negative cases we introduced a new statistical parameter, the detection ratio rate *(DTR). DTR* is expressed as the ratio between the sum of true positive and negative cases with total negative cases, and is use to statistically compare the ability of each test to detect e true case (positive or negative). We analyzed the amniotic liquid provided through amniocentesis and performed in total 80 karyotyping procedures to confirm the total number of the positive screened cases for both first and second trimester cases.

RESULTS AND DISCUSSIONS

The median age for both screened group of pregnant women is 29 years old, lower age 17 years old and higher maternal age was 47 years old. The average age is 29.4 years old for first trimester groupwith confidence interval (CI 28.00 - 30.00)and 29.2 years (CI 28.10 - 30.00) old for the triple test group. The normality for age distribution according to Kolmogorov-Smirnov test is rejected (*Altman DG et al., 1983*). The age statistical characteristics of both screened groups very similar. The frequency distribution is represented in figure 1. The Skewness coefficient has a *p*-value lower than 0.05 which indicates that the age distribution is not symmetric (*Sheskin, 2011*) (table 1).

The total screened pregnant women by means of combined test were 514 cases and the cut off value is 1:250 (p=0.004); out of 514 cases 487 of them were screened negative (confirmed with one of two methods) so they are true negative cases (94.7%); 24 out of 514 were screened as false positive or 4.67% (confirmed with one of two methods); 8 were screened positive for different chromosomal abnormalities (1.6%) and only 3 (three) or 0.6% were screened and confirmed with T21 (table 2). The total screened pregnant women in the second trimester were 487; a numberof 436 out of 487 were screened and confirmed as true negative cases (89.5%); 48 out of 487 were screened and confirmed as false positive (9.85%); 6 or 1.2% were screened and confirmed with other abnormalities and only 3 (three) or 0.6% were screened and confirmed with T21 (table 2). The $p<0.05$ for the kurtosis indicates that both are leptokurtic with opposite peakedness. The dot-line in the figure1 indicates the optimized normality distribution according to Kolmogorov-Smirnov test that represent the number of cases that should be in the corresponding histogram. Consideringall the above mentioned characteristics of both groups have almost the same age characteristics which means that calculated age risk for both groups is the same.

Table 3 shows the weight characteristics of each screened group. In the first trimester group the age interval is from 42-106kg while for the second trimester group is from 46 to 63kg. The weight is important in computing the MoM weight adjusted values and has an impact in the risk assessment but not in the real risk (*T.M Reynolds et al, 2006).* Extreamly higher or low values of weight do have an impact in the risk assesment due to the fact that the biomarkers levels in maternal serum tend to be lower in heavier pregnant women and higher in a lightweighted ones, because they produced by the fetus or fetus-related. In our case the first trimester group has more extrem weight values than the

second group but this has almost none influence in the risk for T21 as the risk calculation has been computed based on MoM weight adjusted values.The weight adjustment reduces the population variability of the markers *(Wald et al., 1992)*. Based on Kolmogorov-Smirnov test the normality in both groups regarding the maternal weight has been rejected ($p<0.05$).

Based on "null" hypothesis"we have calculated the false positive and negative rate of both screening test. The false positive rate (FPR) indicates the measure of the proportion (in %) of unaffected fetus that are screened as positive.In table 1the number of FP cases for the first trimester is 24 out of 487 and for the second group 48 out of 436, which means that for the first trimester prenatal screening the FPR= 4.7% while for the second group it appears much higher(almost twice more), FPR= 9.92%. The sensitivity, that indicates the measure of the proportion of actual positives which are correctly screened as positive, for the combined test carried out based on "null" hypothesis is 100% (CI 30.48%-100.00%)and almost at the same identical value is for the triple test. The high levels of sensivity of both methods is linked with the absence of false negative cases in our population due to not random selection of the all the cases for both tests. Specificity, that indicates the probability that a negative screened pregnant women has not an affected fetus, forthe combined test was found to be at 95.30% (CI 93.09%-96.97%) while for the triple test was at 90.08% (CI 87.07%-92.60%). The differencies at specifity level are strongly linked with the false positive rate (FPR= 1-specifity). Also the confidence interval of the specifity of the triple test seems to be larger than that of the combine test.

The performances between two tests based on the DTR shows clearly the better performance of combined test. For combined test the respective DTR_1= 96.64% at 4% false positive rate, meanwhile for the triple test the respective DTR_2 = 90.07% at 9.92% false positive rate.

The ROC analysis with area under the ROC curve, with standard error and 95% CI can be interpreted as: *a*. the average value of sensitivity for all possible values of specificity; *b*. the average value of specificity for all possible values of sensitivity; and *c*. the probability that a randomly selected individual from the positive group has a test result indicating greater suspicion than that for a randomly chosen individual from the negative group*(Zhou XH et al., 2002)*.In table 4 the ROC area under the curve (AUC) value for combined testis (ROC(AUC)=0.895 (CI 0.864 to0.920) and is at better levels than that of triple test 0.978 (0.965 to 0.991), but both test have a very high value of "discrimination"*(Zweig & Campbell, 1993)*.The significance level or *p*-value is small ($p<0.05$) then it can be concluded that the Area Under the ROC curve is significantly different from 0.5 and that therefore there is evidence that both tests does have a strong ability to distinguish between the negative and positive cases. In table 4 ata fix value for sensitivity fixed at 90% the specificity for triple test is at lower value 89.46 (CI 86.50 to 91.94) than that of combined test 96% (CI 94.1 to 97.7). In figure 3 is shown the sensitivity versus FPR and is expressed in percentage. Is very clear that for both test the ROC curve passes very close to the left-up corner that means a very high ability of the test to distinguish between the affected fetus and the normal ones. When comparing the performance of the test at a fix value of specificity equal to 90% the level of sensitivity remains high in both tests like 100% but the CI is much larger for triple test (CI 29.2 to100) than for combined test (CI 71.5 to100) and under this threshold the sensitivity for triple test is lost (figure 3).

CONCLUSIONS

The performances of first trimester screening test and that of second trimester screening test for Down syndrome in the Albanian pregnant women group are high. The values sensitivity for both test are very high which means that both test are very usefull for the albanian population health. The higher value of false positive rate of second trimester test (triple test) makes this test less preferable than the first trimester one (combined test). The comparison of DTRs' shows a better ability of combined test rather the triple test(DTR_1=96.64%>90.07%=DTR_2)to detect a true case (positive or negative) at lower FPR (4.7%<9.92%). Based on all statistical characteristics computed, we can come to the conclusion that the combined test has a better performance than triple test regarding screening for T21 in the albanian population.We strongly suggest the first trimester test for Down syndrome screening in albanian pregnant women population.

REFERENCES

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

DeLong ER, DeLong DM, Clarke-Pearson DL (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837-845.

Reynolds, T. M., Penney, M. D., Hughes, H., John, R. (1991) The effect of weight correction on risk calculations for Down's syndrome screening. *Ann ClinBiochem***28**: pp. 245–249.

Spencer, K., Spencer, C. E, Power, M., Dawson, C., Nicolaides, K. H. (2003/a). Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemistry in a one-stop clinic: a review of three years prospective experience. *Br J ObstetGynaecol***110**: pp. 281–286.

Spencer, K., Souter, V., Tul, N., Snijders, R., Nicolaides, K. H. (1999) A screening program for trisomy
1 at 10–14 weeks using fetal nuchal translucency, maternal serum free *β*-human chorionic gonadotropin and pregnancy associated plasma protein-A. *Ultrasound ObstetGynecol***13**: pp. 231–237.

Wald, N. J., Cuckle, H. S., Densem, J. W., Kennard, A., Smith, D. (1992) Maternal serum screening for Down's syndrome: The effects of routine ultrasound scan determination of gestational age and adjustment for maternal weight. *Br J ObstetGynaecol***99**: pp. 144–149.

Cleary-Goldman J, Berkowitz RL. First trimester screening for Down syndrome in multiple pregnancies.SeminPerinatol. 2005;29:395–400.

Neveux LM, Palomaki G, Knight GJ, Haddow JE. Multiple marker screening for Down syndrome in twin pregnancies. PrenatDiagn. 1996;16:29–34.

Sheskin DJ (2011) Handbook of parametric and nonparametric statistical procedures. 5th ed. Boca Raton: Chapman & Hall /CRC.

T M Reynolds, G Vranken, J Van Nuete, Weight correction of MoM values: which method? ClinPathol 2006;59:753-758 doi:10.1136/jcp.2005.034280

Zweig MH, Campbell G (1993) *"Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine."* Clinical Chemistry 39:561-577.

Zhou XH, Obuchowski NA, McClish DK (2002). Statistical methods in diagnostic medicine. Wiley-Interscience.

Table 1-Age statistical characteristics of both screened group

Description	**Unit**	**Values**	
Screening test performed		**Combine test**	**Triple test**
Cases screened	[nr.]	514	487
Minimum maternal age	[years]	17	17
Maximum maternal age	[years]	47	47
Age average	[years]	29.4	29.2
Median age	[years]	29.00	29.00
Median with 95% CI	[years]	28.00 - 30.00	28.10-30.00
Skewness		0.4581 (p<0.0001)	0.1417 (p=0.2026)
Kurtosis		0.3805 (p=0.1060)	-0.4877 (p=0.0046)
Kolmogorov-Smirnov test for normal distribution (age)		Rejected normality (p=0.0053)	Rejected normality (p=0.0080)

Table 2-Distribution of cases for combine and triple tests

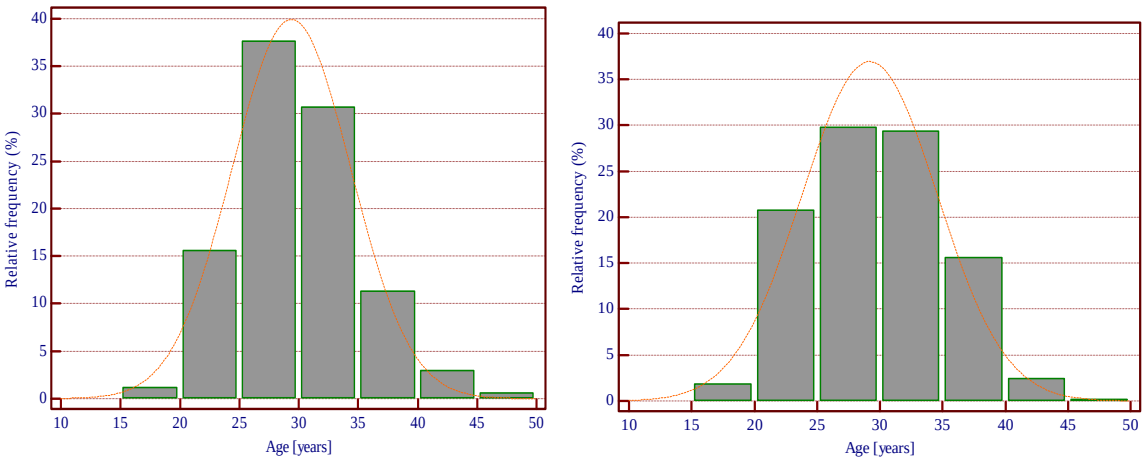
Group	**Number of cases**	
	Combined test	**Triple test**
Total screened pregnant women	514	487
True negative confirmed cases (TN)	487	436
False Positive confirmed cases (FP)	24	48
Confirmed other abnormal cases	8	6
Confirmed Down syndrome cases	3	3
False negative (FN)	0	0

Table 3-Weight statistical characteristics of both screened pregnant women group

Descriptin	**Unit**	**Values**	
Screening test		**Combined test**	**Triple test**
Cases screened	[nr.]	514	487
478	[kg]	42	46
42	[kg]	106	63
110	[kg]	62.00	50
65.00	[kg]	62.94	54.22
65.03	[kg]	61.00-63.00	47.28 deri 61.86
63 deri 65		1.0715 (P<0.0001)	0.1938 (P=0.7813)
0.6845 (P<0.0001)		2.2057 (P=0.0001)	-2.2637 (P=0.0163)
0.7947 (P=0.0056)		Rejected normality (P=0.0001)	Rejected normality (P=0.4134)

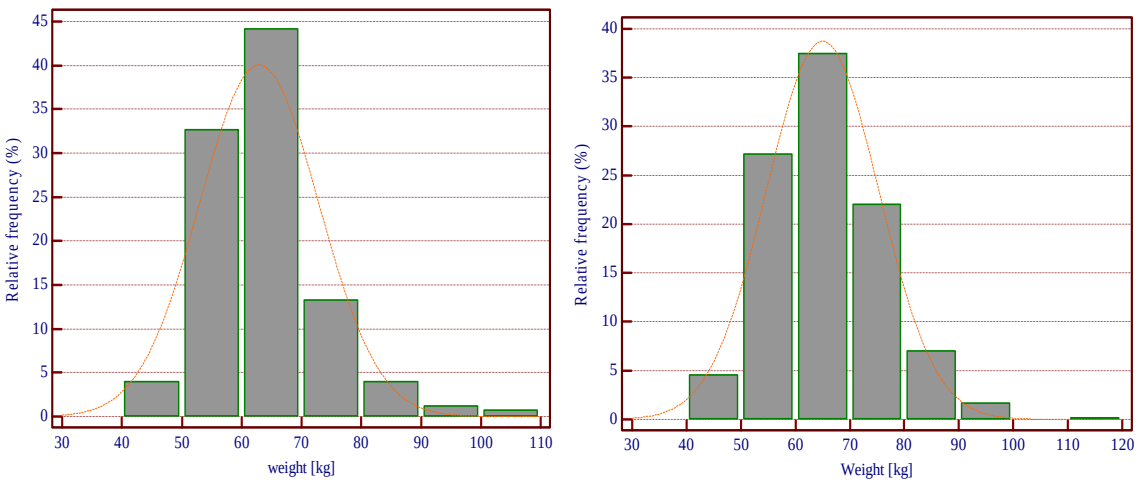
Table 4 Results of ROC analysis for combined and triple test (*DeLong et al*)

Description	**Triple test**	**Combined test**
Total number of cases:	487	513
Positive cases:	3	3
Area Under Curve ROC (AUC) (±1.96 SE)		
ROC (AUC)(CI 95%)	0.895 (0.864 to0.920)	0.978 (0.965 to 0.991)
Standart deviation	0.0140	0.00668
Significance level p (sip=0.5)	<0.0001	<0.0001
Specificity for a fix rate of sensitivity		
Sensitivity [%]	Specificity [%](CI 95%)	
90.00	89.46 (86.50 to 91.94)	96 (94.1 to 97.7)
Sensitivity at fix rate of specificity		
Specificity [%]	Sensitivity [%] (CI 95%)	
90	100 (29.2 to100)	100 (71.5 to100)



A **B**

Figure 1-Histogram of age frequency distributions: **A**. combined test **B**. triple test



A **B**

Figure 2- Histogram of maternal weight distribution: A. combined test B. triple test

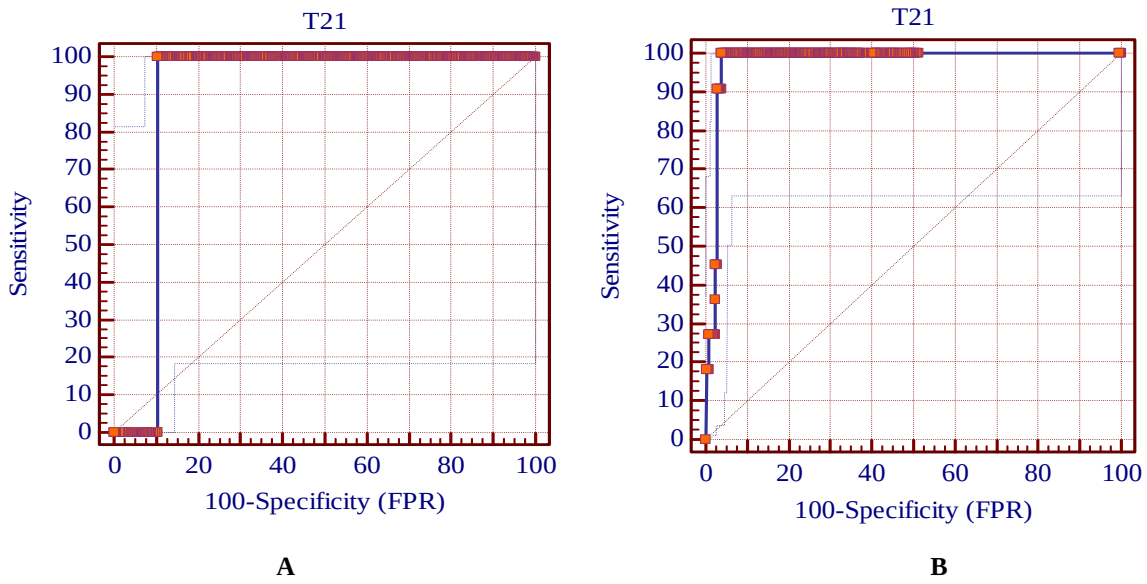


Figure 3 - ROC analysis for Down syndrome: A. combined test; B. triple test