Diagnostic and Interventional Imaging (2016) xxx, xxx—xxx

ORIGINAL ARTICLE /*Genitourinary imaging*

Herman score in prenatal screening for Down syndrome: Can a junior assess a senior?

N. Hamdaoui [a,*], E. Lesieur [a], J. Mancini [b], A. Dabadie [c], V. Bourderionnet [a], H. Pico [c], M. Colombani [c], F. Bretelle [a], G. Gorincour [a,c]

[a] *Prenatal Diagnosis Centre, hôpital de la Timone, 264, rue Saint-Pierre, 13385 Marseille, France*
[b] *Department of Biostatistics, hôpital de la Timone, 264, rue Saint-Pierre, 13385 Marseille, France*
[c] *Department of Pediatric and Prenatal Imaging, hôpital de la Timone, 264, rue Saint-Pierre, 13385 Marseille, France*

KEYWORDS
Pregnancy;
Trisomy 21;
Nuchal translucency;
Herman score

Abstract
Purpose: To compare Herman scores self-assessed prospectively during ultrasound first-trimester screening by a single senior radiologist with 15 years of experience, to those obtained retrospectively by an unexperienced junior radiologist.
Materials and methods: Over a 18-month period, a single senior radiologist measured the nuchal translucency thickness along with calculation of Herman scores. An independent junior radiologist subsequently reviewed and scored the images.
Results: A total of 301 patients were included. The mean Herman score was 8.2 ± 0.9 (SD) for the senior radiologist and 7.8 ± 0.9 (SD) after review by the independent junior radiologist ($P < 0.001$). The scores for caliper position and fetal head position decreased significantly after the independent review. The two criteria on which the two operators disagreed the least were visualization of the nuchal translucency and the distinction between neck and amnios.
Conclusion: Herman score is lower after review by a junior radiologist, without any effect on patient's management and follow-up.

* Corresponding author.
E-mail addresses: hamdaouinaima2611@gmail.com, hamdaoui-naima@hotmail.fr (N. Hamdaoui).

http://dx.doi.org/10.1016/j.diii.2016.03.015
2211-5684/

Please cite this article in press as: Hamdaoui N, et al. Herman score in prenatal screening for Down syndrome: Can a junior assess a senior? Diagnostic and Interventional Imaging (2016), http://dx.doi.org/10.1016/j.diii.2016.03.015

Down syndrome (DS) is defined by the presence of an extra chromosome 21 at karyotype analysis. It is the most frequent chromosomal abnormality with a prevalence of 1 in 700 [1]. It is the leading cause of mental retardation [2]. The diagnosis is based on fetal karyotype analyzed predominantly on fetal cells from amniotic fluid sampled by amniocentesis [3], or more recently on trophoblastic biopsy, following changes brought into the French guidelines [4]. Because of the risk of fetal loss inherent to these invasive techniques, they cannot be offered routinely to all women. Screening consists in targeting patients with an increased risk of chromosomal abnormalities for whom fetal sampling will be considered. The French Health Authorities (Haute Autorité de santé [HAS]) issued in June 2009 recommendations on prenatal screening [5]. Basically, a combined test should be proposed to pregnant women whatever their age. The test should be performed between of 11 + 0 and 13 + 6 week's gestation. The test is based on the age of the mother, as well as ultrasound with nuchal translucency (NT) measurement and biological data (PAPP-A [pregnancy associated plasma protein A] and free-β human chorionic gonadotrophin [free β-hCG]) [1—5]. If the risk of having a child affected by DS is equal or greater than 1:250, invasive fetal sampling is proposed [4]. The threshold of 1:250 provides a sensitivity of 83% for a false positive rate of 3 to 5% [6]. The calculation of the combined risk is only possible if the NT is measured from an ultrasound image of indisputable quality. This requires test standardization and measurement quality control. Technical difficulties can make it impossible to analyze an image and calculate the risk of DS. The Herman score has long been useful to assess objectively the quality of images [5—7]. The training of radiologists, adherence to a quality assurance program to evaluate professional practices and registration into a network of perinatal experts are mandatory [8]. ''Clinical audits'' are proposed to health professionals to evaluate how they measure NT. These remote audits are reviews of images sent via Internet and sent back in the form of a detailed report [9]. However, these ''clinical audits'' are currently not standardized and do not allow the sonographers to self-assess the quality of the images nor to compare their own assessment to that of other reviewers.

The main objective of this study was to compare Herman scores self-assessed prospectively during ultrasound at the first-trimester screening by a single senior operator (SO) who was a radiologist with 15 years of experience, to those obtained retrospectively after review by an independent unexperienced junior operator (UJO) who was a resident in radiology. The secondary objectives were to identify Herman score criteria that varied the most between the two operators.

Materials and methods

Study design

This retrospective study included NT measurements from ultrasound screenings performed during the first trimester of pregnancy in a single imaging department from January 2012 to April 2014, by a single SO with 15 years of experience. The study included all the pregnant women, regardless of their age, who, during the first trimester of pregnancy, underwent ultrasound screening with NT measurement and Herman score.

Exclusion criteria included women with multiple pregnancy, pregnancy scanned before 11 + 0 weeks or after 13 + 6 weeks' gestation (the crown-rump length lower than 45 or greater than 84 mm), and women who had not been assayed for serum markers during the first trimester (PAPP-A and free-β human chorionic gonadotropin).

Data collection

The following data were obtained from all the patients: age, weight, tobacco smoking (yes/no), a previous history of aneuploidy, serum markers assay (PAPP-A and free-β hCG expressed in multiple of median [MoM]), combined risk results, the course of pregnancy and birth (or medical termination of pregnancy [TOP]). The measurement of the NT was expressed in millimeters with precision to the tenth of a millimeter.

Herman score was used to define the quality of the images. The score was based on six criteria. Three ''major'' criteria were rated 0 or 2 depending on whether they were absent or present. These three criteria were: sagittal section, proper placement of the calipers and the visualization of a continuous skin line along the fetal nuchal region. Three criteria were minor criteria and were rated 0 or 1. They corresponded to the size of the area of interest being greater than 75% of the image, amnios visualization along the fetal back and a straight fetal head position. The images were considered usable for the calculation of the combined risk if the score obtained was greater than or equal to 4. When the score was between 4 and 7, the image was deemed satisfactory. The image was considered excellent if the score was between 8 and 9. A total score between 0 and 9 was obtained for each patient for each operator. The sonographer carried out the measurement prospectively and entered the results into Viewpoint software. The images were reviewed subsequently by an UJO who calculated Herman scores blindly unaware of the prospective scores self-assessed by the first operator. The SO worked within a quality assurance program aimed to evaluate professional practice and belonged to a network of experts in perinatology. The UJO was a resident in obstetrics and gynecology with a sole short theoretical training in the analysis of NT images. Finally, an analysis of the number of false-positives (number of fetal sampling who proved normal), of true-positives and false-negatives was made.

Statistical analysis

Data were analyzed using the SPSS software version 20.0 and the package ''psy'' for R 3.1. (weighted kappa). The statistical significance was defined by a P-value < 0.05. Descriptive statistics (numbers/percentages, mean-median ± standard deviation [SD]) were used to summarize the clinical variables. Paired Student t-test was used to compare the mean measurements obtained by the two observers. The weighted kappa coefficient was calculated to evaluate the agreement of the ordinal data obtained by the two observers. The Pearson correlation coefficient was used to determine the relationship between the score of Herman and the weight of the patients.

Please cite this article in press as: Hamdaoui N, et al. Herman score in prenatal screening for Down syndrome: Can a junior assess a senior? Diagnostic and Interventional Imaging (2016), http://dx.doi.org/10.1016/j.diii.2016.03.015