2010

Do race-specific definitions of short long bones improve the detection of Down syndrome on second-trimester genetic sonograms?

Lorie M. Harper

Do Race-Specific Definitions of Short Long Bones Improve the Detection of Down Syndrome on Second-Trimester Genetic Sonograms?

Lorie M. Harper, MD, Diana Gray, MD, Jeffrey Dicke, MD, David M. Stamilio, MD, MSCE, George A. Macones, MD, MSCE, Anthony O. Odibo, MD, MSCE

Objective. The purpose of this study was to determine whether the use of race-specific definitions of short femur and humerus lengths improves Down syndrome detection. ***Methods.*** This was a retrospective cohort study over 16 years. For each self-reported maternal race (white, African American, Hispanic, and Asian), we evaluated the efficiency of Down syndrome detection using published race-specific formulas compared with a standard formula for short femur and humerus lengths (observed versus expected lengths ≤0.91 and ≤0.89, respectively). The sensitivity, specificity, and 95% confidence intervals for each parameter were compared. Screening performance was compared by areas under the receiver operating characteristic curves. ***Results.*** Of 58,710 women, 209 (0.3%) had a diagnosis of a fetus with Down syndrome. Although the race-based formula increased sensitivity in each population, the increase was statistically significant only in the white population, whereas a decrease in specificity was statistically significant in all 4 populations, as denoted by nonoverlapping confidence intervals. The area under the receiver operating characteristic curve for the model using the race-specific definition of short femur length was 0.67 versus 0.65 compared with the standard definition, and for humerus length it was 0.70 versus 0.71. ***Conclusions.*** The use of race-based formulas for the determination of short femur and humerus lengths did not significantly improve the detection rates for Down syndrome. ***Key words:*** Down syndrome; femur length; genetic sonogram; humerus length.

Both femur and humerus lengths have been noted to be shortened in fetuses with trisomy 21 compared with euploid fetuses. A shortened femur length may be defined on the basis of a greater than expected biparietal diameter to bone length ratio,[1,2] a long bone measuring less than the fifth percentile, or a less than expected observed to expected bone length ratio.[3] Despite multiple attempts to refine the definition, the positive predictive value of short femur or humerus length remains less than 10%.[1–3]

Some studies have observed that long bone length varies significantly with maternal race.[4–6] Specifically, African American populations have a longer long bone length compared with white populations, and Asian populations have a shorter long bone length. This observation would suggest that the sensitivity and specificity of a short femur or humerus length is impacted dramatically by race. However, a study using race-based definition of

Received September 1, 2009, from the Department of Obstetrics and Gynecology, Washington University, St Louis, Missouri USA. Revision requested September 22, 2009. Revised manuscript accepted for publication October 15, 2009.

Address correspondence to Lorie M. Harper, MD, Department of Obstetrics and Gynecology, Washington University, 660 S Euclid, Campus Box 8064, St Louis, MO 63110 USA.

E-mail: harperl@wudosis.wustl.edu

expected femur length failed to show an improvement in the detection of Down syndrome in a population of approximately 11,000 patients.[7] Similarly, a study based on a population of 1100 patients did not show a difference in the detection of Down syndrome using race-based definitions of expected humerus length.[8] The conflict between the fact that race affects the observed long bone length but does not affect detection of Down syndrome makes it difficult to counsel women about the impact of race on the association between fetal long bone length and aneuploidy. This study sought to verify prior findings using a larger genetic database with an a priori hypothesis that maternal race does not influence the detection rate of trisomy 21 when using fetal long bone length.

Materials and Methods

This was a retrospective cohort study of patients referred for sonography between 1990 and 2006. Approval was obtained from the center's Institutional Review Board. Sonographic examinations were performed between 16 and 22 weeks. Gestational age was determined by either the last menstrual period if known and concordant with sonography (within 7 days of first-trimester sonography or 14 days of second-trimester sonography) or by the earliest sonogram available when the last menstrual period was unknown or discordant with sonography. Maternal race was self-reported on a questionnaire at the time of sonography. After sonography, the karyotype was confirmed by either prenatal chromosomal analysis or newborn examination. In cases in which newborn examination findings were suspicious for aneuploidy, the karyotype was determined by a newborn blood study.

Previously published definitions of expected femur and humerus lengths were applied to the study population. The regression formulas used in this study are shown in Table 1[3,7,8]; "standard" refers to the formulas generated by Nyberg et al[3] without consideration of race; this formula is referred to as standard in that it has been applied previously to the entire population regardless of race. Descriptive statistics were used when appropriate, and a Student *t* test was used for continuous variables. $P < .05$ was considered statistically significant. The specificity, sensitivity, and 95% confidence intervals of each formula to detect Down syndrome were determined. Statistical significance of the test characteristics were assessed by nonoverlapping confidence intervals. All statistical analyses were performed with Stata version 8.0 software (StataCorp, College Station, TX).

Results

During the study period, 62,111 patients underwent genetic sonography at Washington University, and in this population, a total of 61,185 femur lengths and 43,332 humerus lengths were measured. Of these patients, 41,987 were white, 14,301 were African American, 744 were Hispanic, 1678 were Asian, and 3401 were listed as "other," which included Asian Pacific, American Indian, and Arabic, for which we had insufficient numbers for analysis. A total of 209 cases of Down syndrome occurred in the population: 173 in white patients, 23 in African American patients, 7 in Hispanic patients, and 6 in Asian patients. The most common indications for obtaining sonograms were routine, maternal age of 35 years or older, and abnormal serum screen results. African American patients had a higher rate of routine screening as an indication for sonography, whereas white and Asian patients had a slightly higher frequency of maternal age of 35 years or older as the indication (Table 2).

Table 1. Formulas Used for Expected Femur and Humerus Lengths

Measurement	Formula
Femur length	
Standard[3]	–0.966 + 0.866 × BPD
White[7]	–0.802 + 0.857 × BPD
African American[7]	–0.859 + 0.871 × BPD
Hispanic[7]	–0.851 + 0.869 × BPD
Asian[7]	–0.891 + 0.877 × BPD
Humerus length	
Standard[3]	–0.884 + 0.834 × BPD
White[8]	–7.0058 + 0.7995 × BPD
African American[8]	–6.1769 + 0.7995 × BPD
Hispanic[8]	–6.8024 + 0.7995 × BPD
Asian[8]	–8.0761 + 0.7995 × BPD

BPD indicates biparietal diameter.

Table 2. Population Characteristics

Characteristic	White (n = 41,987)	African American (n = 14,301)	Hispanic (n = 744)	Asian (n = 1,678)
Maternal age, y[a]	31.9 ± 5.9	26.6 ± 6.8[b]	30.0 ± 6.7[b]	32.6 ± 5.1[b]
Down syndrome, n (%)	173 (0.41)	23 (0.16)[b]	7 (0.94)[b]	6 (0.36)
Indication for sonography, n (%)				
Routine	11,900 (28.3)	7,644 (53.5)[b]	324 (43.5)[b]	647 (38.9)[b]
Advanced maternal age	15,148 (36.1)	2,170 (15.2)[b]	191 (25.7)[b]	563 (33.6)[b]
Abnormal serum screen results	3,590 (8.6)	823 (5.8)[b]	130 (7.7)	56 (7.5)

[a]Values are mean ± SD.
[b]Statistically significant (*P* < .05) when compared with the white population.

African American fetuses had a longer femur length than white fetuses for every gestational age (Table 3). At almost every gestational age, white fetuses had a femur length similar to that of both Hispanic and Asian fetuses. When examining humerus length, African American fetuses again had a longer bone length than white fetuses. For most gestational ages, Hispanic fetuses also had a longer humerus length than white fetuses. With the exception of the gestational age range of 20 to 20.9 weeks, Asian and white fetuses had similar humerus lengths.

When the standard formula for observed versus expected femur length was applied to the entire population, the sensitivity for detecting trisomy 21 was 46.4% (95% confidence interval,

Table 3. Average Femur and Humerus Lengths

Gestational Age, wk	Measurement, cm			
	White	African American	Hispanic	Asian
Femur length				
15–15.9	1.92 ± 0.19	1.97 ± 0.21 *P* < .01	1.93 ± 0.20 *P* = .82	1.91 ± 0.20 *P* = .56
16–16.9	2.23 ± 0.20	2.26 ± 0.22 *P* < .01	2.18 ± 0.20 *P* = .45	2.21 ± 0.25 *P* = .33
17–17.9	2.52 ± 0.20	2.56 ± 0.20 *P* < .01	2.51 ± 0.23 *P* = .86	2.51 ± 0.20 *P* = .61
18–18.9	2.86 ± 0.20	2.90 ± 0.19 *P* < .01	2.84 ± 0.20 *P* = .21	2.84 ± 0.17 *P* = .09
19–19.9	3.13 ± 0.20	3.20 ± 0.19 *P* < .01	3.16 ± 0.17 *P* = .15	3.13 ± 0.21 *P* = .88
20–20.9	3.39 ± 0.21	3.44 ± 0.20 *P* < .01	3.39 ± 0.19 *P* = .77	3.36 ± 0.19 *P* = .03
21–22	3.67 ± 0.23	3.74 ± 0.22 *P* < .01	3.64 ± 0.23 *P* = .14	3.64 ± 0.18 *P* = .09
Humerus length				
15–15.9	1.88 ± 0.21	1.90 ± 0.22 *P* = .14	1.88 ± 0.22 *P* = .88	1.86 ± 0.20 *P* = .48
16–16.9	2.16 ± 0.20	2.21 ± 0.22 *P* < .01	2.21 ± 0.22 *P* < .01	2.11 ± 0.25 *P* = .05
17–17.9	2.41 ± 0.21	2.48 ± 0.20 *P* < .01	2.48 ± 0.20 *P* < .01	2.41 ± 0.20 *P* = .89
18–18.9	2.74 ± 0.21	2.81 ± 0.23 *P* < .01	2.81 ± 0.23 *P* < .01	2.73 ± 0.19 *P* = .38
19–19.9	3.00 ± 0.23	3.10 ± 0.21 *P* < .01	3.08 ± 0.21 *P* < .01	2.98 ± 0.21 *P* = .17
20–20.9	3.23 ± 0.22	3.29 ± 0.23 *P* < .01	3.29 ± 0.23 *P* < .01	3.18 ± 0.20 *P* < .01
21–22	3.48 ± 0.24	3.55 ± 0.23 *P* < .01	3.68 ± 0.28 *P* = .07	3.47 ± 0.22 *P* = .46

Values are mean ± SD.

39.5%–53.4%). When the standard formula was applied to each individual race, the sensitivity ranged between 40.0% and 47.3%. When the race-specific formulas were applied to the population, the sensitivity for Down syndrome detection apparently increased to a 60.0% to 71.9% range; however, the only population in which the confidence intervals for sensitivity did not overlap was the white population, indicating that this was a statistically significant improvement only in the white population. In the African American, Hispanic, and Asian populations, use of race-based formulas did not result in a statistically significant increase in the sensitivity.

The specificity of the standard formula applied to the entire population was 84.5% (84.2%–84.85; Table 4). Applied to each population, the specificity of the standard formula varied between 76.8% and 89.6%. When the race-based formulas were used, the specificity ranged between 50.6% and 73.0% for each population. This decrease was statistically significant because the confidence intervals for specificity did not overlap in any population. The area under the receiver operating characteristic curve for the model using the race-specific definition of short femur length was 0.67 compared with 0.65 for the standard definition.

A similar pattern was seen in the formulas for short humerus length (Table 4). The standard formula applied to the entire population resulted in sensitivity of 72.1% (64.1%–79.2%). Applied to each race individually, the sensitivity ranged between 42.9% and 77.1%. When race-specific formulas were used, the sensitivity varied between 83.3% and 100.0%. Again, the confidence intervals for the standard formula and race-specific formulas overlapped in every population except the white population. The specificity of the standard formula applied to the entire population was 70.1% (69.7%–70.6%); applied to each race, the standard formula produced specificity between 60.9% and 77.9%. The race-specific formulas for short humerus length resulted in decreased specificity, ranging from 42.2% to 62.5%. The confidence intervals for the specificity did not overlap, indicating a statistically significant decrease in specificity. The area under the receiver operating characteristic curve for the model using the race-specific definition of short humerus length was 0.70 versus 0.71 for the standard definition.

Discussion

In our cohort of almost 60,000 patients, the use of race-specific definitions for expected femur length did not improve the sensitivity of the test and resulted in decreased specificity for the detection of Down syndrome, corroborating prior published reports by Borgida et al[7] and Mastrobattista et al.[8] Although the race-based formulas resulted in an apparent trend toward increased sensitivity, the statistically significant decrease in specificity mitigates any benefits that might be attained by using the race-based formulas.

Our findings also confirm the findings of prior reports that differences in femur and humerus lengths exist between races.[4–6] However, although these differences are statistically significant, they are small differences, possibly explaining why

Table 4. Sensitivity and Specificity of Race-Based Definitions of Short Femur and Humerus Lengths Compared With Universal Definitions

	Femur Length				Humerus Length			
	Sensitivity, %		Specificity, %		Sensitivity, %		Specificity, %	
	Standard	Race Specific	Standard	Race Specific	Standard	Race Specific	Standard	Race Specific
All	46.4 (39.5–53.4)	NA	84.5 (84.2 – 84.8)	NA	72.1 (64.1–79.2)	NA	70.1 (69.7–70.6)	NA
White	47.3 (39.5–55.2)	71.9 (64.4–78.5)	83.1 (82.7–83.5)	61.3 (60.8–61.8)	77.1 (68.4–84.3)	91.5 (85.0–95.9)	67.8 (67.2–68.3)	47.1 (46.6–47.7)
African American	40.9 (20.7–63.6)	63.6 (40.7–82.8)	89.6 (89.1–90.1)	73.0 (72.3–73.8)	42.9 (17.7–71.1)	85.7 (57.2–98.2)	77.9 (77.1–78.7)	62.5 (61.6–63.4)
Hispanic	42.9 (9.9–81.6)	71.4 (29.0–96.3)	83.1 (80.2–85.8)	61.3 (57.6–64.8)	83.3 (35.9–99.6)	83.3 (35.9–99.6)	66.8 (63.1–70.4)	49.3 (45.4–53.1)
Asian	40.0 (5.2–85.3)	60.0 (14.7–94.7)	76.8 (74.7–78.9)	50.6 (48.1–53.0)	50.0 (6.7–93.2)	100.0 (39.7–100.0)[a]	60.9 (58.2–63.5)	42.2 (39.5–44.8)

Values in parentheses are 95% confidence intervals. NA indicates not applicable.

[a]One-sided, 97.5% confidence interval.

accounting for these differences does not improve the prenatal detection of Down syndrome. One difference between our findings and prior reports is that most reports found that Asian fetuses had a shorter long bone length when compared with white fetuses, whereas in our population, Asian and white fetuses had similar femur and humerus lengths.

One of the limitations of this study was the use of previously published formulas. The differences in the regression lines for expected long bone length developed by individual institutions has led to the suggestion that Down syndrome detection would be enhanced by each institution's developing its own regression curves.[9] However, differences in the racial composition at each institution may have explained why individual institutions developed slightly different regression curves. Furthermore, most sonography units do not have large enough patient populations to develop institution-specific, race-specific regression curves and would need to use previously published formulas if they were to use race-specific formulas. We therefore elected to use previously published formulas because that would be in keeping with common practice.

We also used self-reported maternal race, in keeping with prior published reports. This does not take into account paternal race, which may also affect femur and humerus lengths.

The size of the study was its main strength, with more than 60,000 femur lengths and more than 40,000 humerus lengths, as well as 209 cases of trisomy 21. This population also included a large African American subset. We had smaller numbers of Asian and Hispanic patients in this population; however, these numbers compared with the numbers included in the studies by Borgida et al[7] and Mastrobattista et al.[8]

In summary, in this large retrospective cohort, the use of race-specific definitions of short femur and humerus lengths resulted in no improvement in sensitivity and decreased specificity when compared with a standard definition. At this point, the use of race-specific regression formulas to determine short femur and humerus lengths does not improve the detection of Down syndrome on second-trimester genetic sonograms.

References

1. Rodis JF, Vintzileos AM, Fleming AD, et al. Comparison of humerus length with femur length in fetuses with down syndrome. Am J Obstet Gynecol 1991; 165:1051–1056.
2. Dicke JM, Gray DL, Songster GS, Crane JP. Fetal biometry as a screening tool for the detection of chromosomally abnormal pregnancies. Obstet Gynecol 1989; 74:726–729.
3. Nyberg DA, Resta RG, Luthy DA, Hickok DE, Williams MA. Humerus and femur length shortening in the detection of Down's syndrome. Am J Obstet Gynecol 1993; 168:534–538.
4. Shipp TD, Bromley B, Mascola M, Benacerraf B. Variation in fetal femur length with respect to maternal race. J Ultrasound Med 2001; 20:141–144.
5. Kovac CM, Brown JA, Apodaca CC, et al. Maternal ethnicity and variation of fetal femur length calculations when screening for Down syndrome. J Ultrasound Med 2002; 21:719–722.
6. Zelop CM, Borgida AF, Egan JF. Variation of fetal humeral length in second-trimester fetuses according to race and ethnicity. J Ultrasound Med 2003; 22:691–693.
7. Borgida AF, Zelop C, Deroche M, Bolnick A, Egan JF. Down syndrome screening using race-specific femur length. Am J Obstet Gynecol 2003; 189:977–979.
8. Mastrobattista JM, Pschirrer ER, Hamrick MA, et al. Humerus length evaluation in different ethnic groups. J Ultrasound Med 2004; 23:227–231.
9. Vergani P, Locatelli A, Piccoli MG, et al. Critical reappraisal of the utility of sonographic fetal femur length in the prediction of trisomy 21. Prenat Diagn 2000; 20:210–214.