ETHNIC GROUP SPECIFIC NORMATIVE CURVE FOR FETAL FEMUR LENGTH MEASUREMENTS IN UTERO

By

Christine F. Blake, M.D.

A THESIS

Presented to the Department of Public Health and Preventive Medicine

of the Oregon Health Sciences University

in partial fulfillment of the requirements for the degree of

Master in Public Health

April 2004

School of Medicine
Oregon Health and Science University

CERTIFICATE OF APPROVAL

This is to certify that the M.P.H. thesis of

Christine F. Blake, M.D.

has been approved

[redacted]

Kenneth E. James, Ph.D., Chair, Thesis Committee

[redacted]

Donald F. Austin, M.D., M.P.H., Member, Thesis Committee

[redacted]

Jeffrey T. Jensen, M.D., M.P.H., Member, Thesis Committee

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CI- confidence interval

cm- centimeters

CPT- Current Procedural Terminology

EBC- Electronic Birth Certificate

EGA- estimated gestational age

FL- femur length

GA- gestational age

Kg- kilogram

M^3- cubic meters

mm- millimeters

OHSU- Oregon Health and Science University

SD- standard deviation

ACKNOWLEDGMENTS

I would like to thank the following departments for their assistance in providing the data for this project:

Department of Radiology
Oregon Health and Science University

Department of Obstetrics and Gynecology
Oregon Health and Science University

ABSTRACT

Currently, over 2 million obstetrical ultrasounds are done each year in the United States (Martin et al, 2002). In Healthy People 2010, one of the goals is a reduction in racial and ethnic differences in health status in the United States. This was a retrospective cohort study designed to determine whether the use of current Caucasian ultrasound growth curves result in a higher than expected prevalence of abnormally short femur lengths, when they are applied to a Hispanic population. Ultrasound examinations in 231 Hispanic women over an 18-month time frame were used to generate a Hispanic femur length normative curve. The prevalence of short femur diagnoses (less than 5th percentile) for this Hispanic curve as well as to the standard reference (Hadlock, 1984) were calculated after application to a separate 6-month validation cohort. For this validation sample the rate of short femur diagnoses for the standard reference was 12.5% and for the new Hispanic normative curve was 8.0% . These values were compared using McNemar's test yielding an insignificant p value of 0.125. The potential cost savings for the use of the new Hispanic curve were analyzed. If the excess false positive rate of 4.5% is a valid estimation and each diagnosis of short femur results in a single follow up ultrasound examination, for every 1000 Hispanic patients undergoing ultrasound, the application of the new Hispanic curve could save $5400 by reducing the number of false positive short femur diagnoses. Since dwarfism is so rare, only 1-3 cases per 10,000 persons, adequate assessment of the ascertainment of dwarfism for the new Hispanic model could not be performed. Until this remaining issue can be resolved, it would be premature to recommend replacement of the current standard with this new model for Hispanic patients undergoing obstetrical ultrasound examination.

INTRODUCTION

In the 21st century, the majority of pregnancies in the United States and other developed nations will have at least one ultrasound examination. In 1999, the most recent year for which comprehensive data are available for the United States, this translated into over 2.5 million ultrasound examinations (Martin et al, 2002). The typical reasons for these examinations are confirmation of expected delivery date, normal fetal growth, or normal fetal anatomy. In a typical obstetrical ultrasound evaluation the following are usually assessed: architecture of uterus and placenta, the fetal number and position, fetal biometric information (the size of the fetal head, abdomen and femur (thigh bone)), the presence of cardiac and fetal movement, and an assessment of the internal anatomy of the fetus.

If the ultrasound examination reveals an abnormality, follow-up examinations to confirm the finding and assess for progression of the abnormality are generally recommended. Careful training of personnel and establishment of appropriate norms are critical in avoiding unnecessary false positive diagnoses of abnormalities. False positive abnormal diagnoses are a source of great stress to the pregnant patient and her family, as well as a source of unnecessary utilization of health care resources.

In Healthy People 2010, one of the goals is a reduction in racial and ethnic differences in health status in the United States. As with many areas of health care, distinct variation exists in pregnancy outcome in the United States by racial and ethnic divisions. The extent to which these differences are due to differential access to health care, socioeconomic status, cultural beliefs, or intrinsic differences in biology, remains to be discovered.

In obstetrical ultrasound, many of the normative scales are derived from ethnically homogenous populations. A normative curve developed for a middle class Caucasian population may not apply to a more ethnically heterogeneous population. In order to correctly identify abnormal growth within a population, one must first have accurately defined normal growth (Otto and Platt, 1991). The current fetal biometric ultrasound norms used at Oregon Health and Science University (OHSU) were derived over a decade ago, from an essentially Caucasian population (Hadlock ,1984). Using these established norms, each fetus may be given a percentile ranking for size, based on established gestational age. Most of the time, the use of these norms seems to be adequate in a multi-ethnic population. However, when composite or individual biometric results outside the 5th or 95 percentile are found, follow-up ultrasound testing will usually be recommended.

The idea for this research project arose from the anecdotal observation that among the Hispanic population at OHSU, the abnormal diagnosis of femur length less than fifth percentile seems to occur significantly more often than 5 percent of the time. These observations lead to my hypothesis that fetal femur length measurements for the Hispanic population differ from the standard reference measurements. The purpose of this research is to derive ethnic-specific norms for fetal femur length measurements for a Hispanic population with the goal of reducing unnecessary follow-up ultrasound testing within these populations.

BACKGROUND

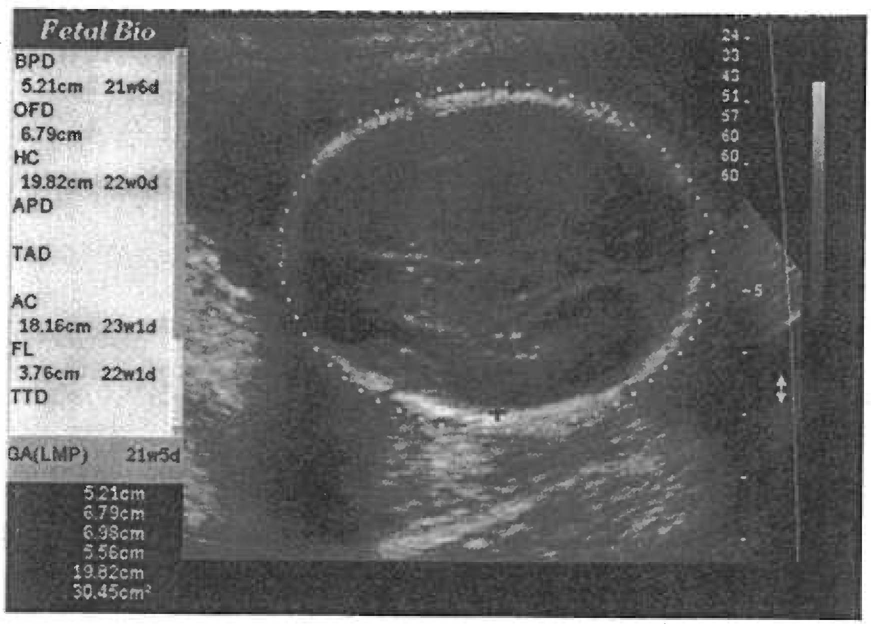
Obstetrical Ultrasound

In the United States, the majority of pregnancies receive at least one ultrasound for assessment of fetal size and anatomy. Measurement of fetal head size, abdomen size and femur length are routine during these examinations and are referred to as fetal biometry. Normative tables of fetal biometric measurements and gestational age are used in these examinations. There are several well-known reference tables for each of the standard measurements (Hadlock, 1984; Jeanty, 1984; Doubilet, 1993). A combination of these biometric measurements is then usually used to provide an estimate of fetal weight and a composite estimate of fetal size and/or age.

In general, these normative curve studies involve ultrasound examination of well-dated pregnancies. The biometric measurements and known gestational age are then used to create linear regression models using the method of least squares to identify the best fit. Many different models have been constructed using different measurable fetal parts, however, the most commonly used are the head size, abdomen size and femur length.

The head size is measured in two ways, at the same transverse anatomic plane (Figure 1). The first is called the biparietal diameter, and is the distance across the head at the level of the thalami. The second is the head circumference, which is the distance around the head at this same level. Both measurements are used because the bones of the fetal skull are not fused and head shape may vary based on fetal position and presentation.

FIGURE 1: Ultrasound Image of Fetal Head

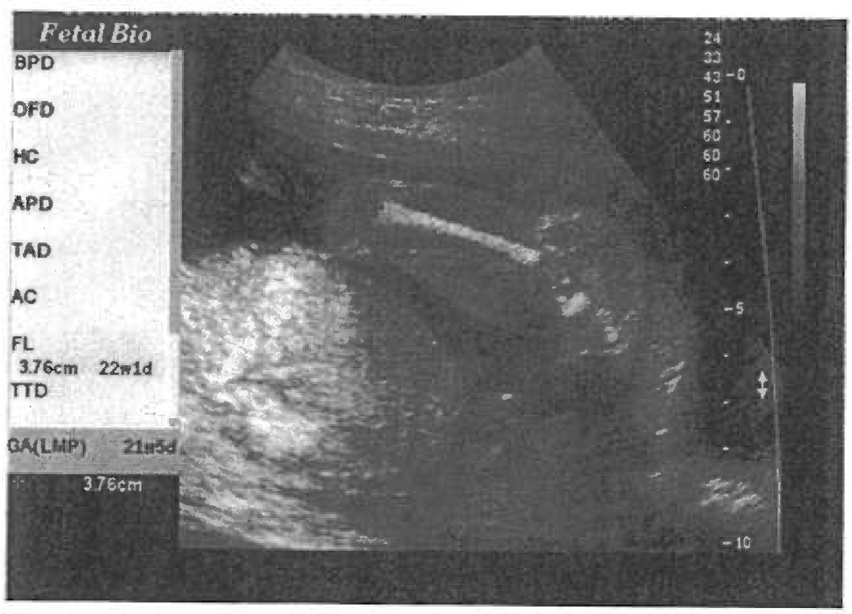


The fetal abdomen is assessed with a single measurement called the abdominal circumference (Figure 2). This image is taken in a transverse plane at the level of the stomach and umbilical vein.

FIGURE 2: Ultrasound Image of Fetal Abdomen

The fetal femur length is assessed ideally in a horizontal position relative to the ultrasound probe (Figure 3).

FIGURE 3: Ultrasound Image of Fetal Femur



With early ultrasound equipment, the practitioner would take the measurements and then utilize a table or graph to convert the biometric measurement to gestational age equivalent. For most ultrasound equipment in current use, the normative tables have been preloaded into the equipment so that this conversion of size to gestational age is instantaneous as depicted above in the tables on the left of Figures 1, 2 and 3.

Ethnic Variation in Biometric Parameters

There is a large body of literature dealing with racial and ethnic variation in pregnancy outcomes such as rates of low birth weight and infant mortality. The attribution of these differences to racial or ethnic categories is not clear-cut. The increased risks of adverse outcome within certain groups, particularly blacks, persist in some studies even after correction for other known risk factors such as education level, insurance status, socioeconomic status, age, and smoking (Krieger, 1993).

An example of where application of typical Caucasian norms to an ethnic population may not be appropriate can be found in a study done in England on a low socioeconomic status population where anthropometric/biometric measurements of Caucasian and Indian infants were compared (Moore et al, 1995). They found that 73% of the boys of Indian mothers fell below the 50th percentile of the standard Manchester curve (Yudkin et al, 1987) for abdominal circumference measurements, while only 52% of the boys of the English mothers were below this line. Furthermore, comparison of the ponderal indices (kg/m^3 or ratio of weight to body volume) of these infants demonstrated that this difference was not due to a reduction in the deposition of body fat, but rather due to differential distribution of body fat (Moore et al, 1995).

In a similar study, looking at ethnic groups within a middle class Kaiser Permanente population, it was noted that the mean birth weights of both Hispanic and Asian babies were slightly less than white babies (Shiono et al, 1986). The mean birth weight for white babies was 3,480 grams in this study, and the mean birthweights for Hispanic babies and Asian babies were 105 grams and 210 grams less than mean birth weight for white babies respectively (Shiono et al, 1986).

These data would seem to imply an effect of ethnicity and race on commonly used biometric parameters, particularly infant birthweight. The assessment of the effects of race or ethnicity on pregnancy norms is an area of epidemiologic research complicated by multiple confounders and the largely retrospective nature of the majority of the data. In a recent study on the use of a diagnosis of small for gestational age (weight below 10^{th} percentile for age) to predict neonatal morbidity and mortality, the authors found that a regression equation which included certain maternal and infant characteristics, such as maternal height, prepregnancy weight, age, parity, weight gain, and race, improved the diagnosis of infants at risk for increased perinatal morbidity and mortality (Sciscione, Gorman, and Callan, 1996). An accurate understanding of the effects of these confounding variables, such as race or ethnicity, and how they may modify baseline risk, may assist in the process of risk assessment. Appropriate diagnosis and risk assessment will assist with more appropriate resource utilization, as well as potentially reduced overall costs by reducing unnecessary follow up ultrasound examinations.

While considerable attention has been paid to the ethnic differences in undesirable pregnancy outcomes, the subject of fetal anthropometric or biometric differences between infants of different ethnic groups has received little attention. Certainly, it is well known that for children and adults the normal growth curves and the mean and range of height vary among ethnic groups (Rimoin, Borochowitz, and Horton, 1986). Two studies in recent years have addressed ethnic differences in fetal ultrasound biometry. The first was a prospective cross-sectional study comparing the fetuses of pregnant women of native Belgian origin to Moroccan and Turkish immigrants (Jacquemyn et al, 2000). This study demonstrated statistically significant differences in the generated polynomial regression coefficients between the ethnic groups for head circumference, abdominal circumference, and femur length as a function of gestational age. The second study was a prospective comparison of body proportions, specifically a ratio of fetal femur length to neonatal crown-heel length (height) among fetuses of women of Malay, Chinese or Indian descent (Lim et al, 2000). This study demonstrated a significant difference in body proportions of the Indian population from the Chinese and Malay populations.

Two other studies have addressed the racial differences between fetal biometric ultrasound norms in black versus white populations (Davis et al, 1993; Hill et al, 1989). Both of these studies found that the mean lengths of the long bones of the arms and legs were consistently greater in black versus white babies throughout gestation. They concluded, however, that these differences were small and that overall the use of the standard norms (based on a Caucasian population) did not affect the estimation of gestational age significantly.

In general, a bone length greater than the 95th percentile is not a cause for concern, however a bone length below the fifth percentile may be an indicator of suboptimal growth or fetal structural anomaly. Because an unusually long leg does not normally prompt follow up assessment, the authors of the previous study (Davis et al, 1993) did not address the economic repercussions that might result from the use of an inappropriate normative curve. In these instances in which the of femur length is less than the 5th percentile, further ultrasound studies would generally be ordered to be sure the infant did not have a growth abnormality or a skeletal dysplasia, resulting in increased costs. Certainly, at OHSU, the anecdotal observation is that among Hispanics, femur length measurements fall below the fifth percentile much more than 5% of the time and no published norms exist for the Hispanic ethnic population.

SPECIFIC AIMS

The overall goal of this project is to determine whether the use of current Caucasian ultrasound growth curves result in a higher than expected prevalence of abnormally short femur lengths, as assessed by antenatal ultrasound, when they are applied to a Hispanic population. The project focuses on femur length because anecdotal observations have not revealed an unusually high level of abnormal head or abdominal size. In order to accomplish this goal, the project has three specific aims:

1. To collect ultrasound biometric data on femur length measurements, for all pregnant patients delivering between January 1, 1995 and June 30, 1996, and to use these data to construct a separate regression curve of femur length versus gestational age for an OHSU Hispanic population. The data set for this analysis is called the exploratory or model development sample.
2. To apply the newly generated ethnic-group specific growth curve to a separate July 1, 1996 to December 31, 1996 validation sample of delivered Hispanic patients to estimate the incidence of short femur length (less than fifth percentile). This incidence is compared to the incidence using the Hadlock growth curves for a Caucasian population.
3. To use the information obtained in aim two and the standard Medicare ultrasound reimbursement information to construct a cost savings analysis.

RESEARCH DESIGN

The format of the project overall is that of a retrospective cohort study of all women delivering infants at the Oregon Health and Science University

(OHSU), over a two-year time frame. Pregnant women who initiated prenatal care at OHSU or one of its satellite clinics, underwent obstetrical ultrasound during the pregnancy, and subsequently delivered at OHSU comprised the cohort that was studied.

In addition to the specific information needed for generation of the ultrasound normative curves, general and demographic information regarding pregnancy and delivery of all women delivering between January 1, 1995 and December 31, 1996 was abstracted from the computerized birth registry at OHSU. These data included maternal date and time of delivery, age, gestational age at delivery, birth weight, weight gain of the mother, time of care initiation, mother's education level, mother's tobacco use and marital status.

MATERIALS AND METHODS

Data Collection and Processing

Data were obtained from two OHSU databases: the obstetrical ultrasound database, which contained data from March 1994 through December 1996 and the delivery database consisting of the Labor and Delivery Logbook and the Electronic Birth Certificate (EBC).

Data from the OHSU ultrasound database, a comma delimited file, was exported into Excel 2000 for initial editing and manipulation. Patients were excluded if for multiple gestations, anembryonic gestation, early exams (e.g. <13 weeks) where a femur length was not calculated, and uncertain date of last menstrual period. Patients were also excluded if their ultrasound examination was inconsistent with last menstrual period dating. In obstetrical ultrasound, previous studies have demonstrated that the standard deviation and desired confidence interval for any biometric measurement progressively increase with advancing gestational age due to progressive expression of inherent normal biologic variability by the fetus as it grows. This trend has been noted in virtually all studies of fetal biometric data obtained by ultrasound. Some authors have actually used the biometric data from well dated pregnancies to create "backwards" regressions of EGA on the biometric measurements to determine the precision with which gestational age can be estimated by ultrasound(Hadlock et al; 1984; Gabbe et al, 1990). Based on composites of several studies, the ultrasound examinations have traditionally been divided into four groups of 13-20 weeks, 21-24 weeks, 25-32 weeks, and 32 weeks to term (Gabbe et al, 1990). For each grouping, the upper limit of the clinically accepted confidence interval is set as in Table 1. These groupings represent clinical distinctions based on multiple previous studies (Hadlock et al, 1984; Gabbe et al, 1990). By standard clinical practice, if the average or weighted average of the four measurements fell outside the predicted range, the dating information would be reviewed, and the due date possibly changed.

TABLE 1: Clinically Accepted Margin of Error for Estimation of Menstrual Age from Ultrasound Biometric Parameters throughout Gestation

Weeks gestation	Upper limit of CI
13-20	10 days
21-24	13 days
25-32	16 days
32- 42	21 days

For our data, the examinations within age each grouping were evaluated, and if the composite gestational age by the standard four measurements of biparietal diameter, head circumference, abdominal circumference, and femur length fell outside the upper limit estimation, the examination was deemed incompatible with dating and was eliminated from the database. This was done so that all examinations used to generate the growth regression curve would be from well-dated pregnancies.

Data from the Labor and Delivery Log book for the period January 1, 1995 though December 31, 1996 was merged with the EBC using the mother's name and date and time of delivery. This merged database contained the maternal and infant medical record numbers and the maternal country of origin and ethnic background. Finally, the ultrasound and delivery databases were merged using the newborn medical record identification numbers and the ethnic data to provide a database of Hispanic women delivering at OHSU who had ultrasound examinations performed during their pregnancies. This merged database was then divided into two time periods: mothers delivering between January 1, 1995 and June 30, 1996, containing the exploratory/model building sample, and mothers delivering between July 1, 1996 and December 31, 1996, containing the validation sample.

Data Analysis

Pertinent demographic information was compared between the subgroup of Hispanic women with ultrasounds, the subgroup of Hispanic women without ultrasounds and the delivering population overall to determine whether demographic differences between the groups might account for observed differences in the incidence of short femur length. Continuous variables were compared with the student t-test and categorical variables were compared using chi-square contingency tables.

Data for women delivering from January 1, 1995 to June 30, 1996 were used to fulfill aim 1 to construct the ethnic group specific normative curve for femur length in the Hispanic population. This process involved several steps: First the femur lengths were plotted against known gestational age to assess the nature of the relationship between these measurements. Then, alternative regression equation models (linear, quadratic, cubic) relating gestational age to femur length were estimated. In comparing these equations, we sought the simplest model in which all estimated coefficients were significantly different from zero and the residual mean square was minimized.

The data from Hispanic patients delivering from July 1, 1996 to December 31, 1996 were used to accomplish aim 2. Using this 6-month validation cohort, we compared the incidence of less than fifth percentile femur length in our Hispanic population, using the OHSU Hispanic curve generated in aim 1, versus the incidence by the Hadlock standard growth curve used at OHSU and other hospitals in the United States. McNemar's (Siegel S, Castellian NJ, 1988) test was used to compare the incidences of short femur length estimated by the two growth models.

The information collected in aims 1 and 2 were used to construct a cost savings analysis, evaluating the reduction in follow up costs generated by using an ethnic group specific normative curve for the Hispanic population. The Medicare fee schedule for ultrasound was used to attach costs to the ultrasounds saved by using the ethnic-specific growth curves.

SPSS 11.0 for Windows was used to compare the demographic variables, construct and compare the growth curves and perform the cost analysis.

Patient Population

Clearly, due to the unique nature of a pregnant population, all study subjects will be female. The population of delivered infants was not gender restricted. The ethnic and minority group make up of the population was diverse, though highly Hispanic dominated.

Human Subjects

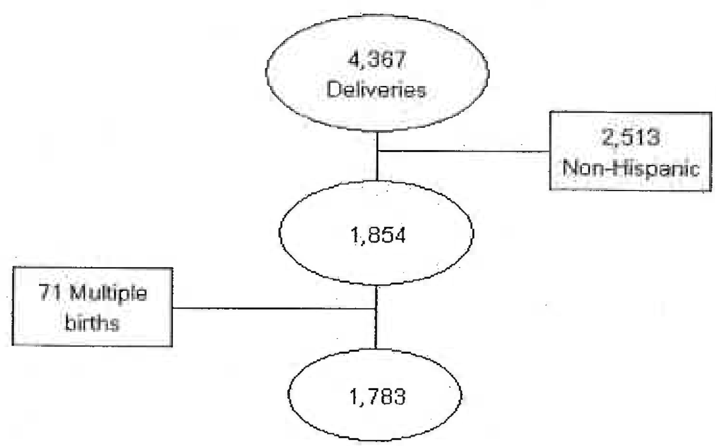
Approval for this retrospective research project was obtained from the Oregon Health Sciences Institutional Review Board. IRB approval was necessary because identifying information from the three original data sources needed to be retained until the information for each mother-infant pair could be merged. After merging of the database information, each mother-infant pair was assigned a unique subject number; and thereafter the information was recorded by the investigator in such a manner that subjects could not be identified. In order to preserve patient confidentiality, access to the databases was password protected.

RESULTS

Data from the obstetrical ultrasound database originally consisted of 10,018 examinations. Figure 4 illustrates how ultrasound examinations inappropriate to use in creating the regression equation were sequentially eliminated from the database. Exclusions were for multiple gestation (1,045), anembryonic gestation (47), femur length not calculated (1,511), menstrual dating missing (4,739), and ill-dated pregnancy (642), leaving a total of 2,034 pregnancies to match to delivered Hispanic mothers.

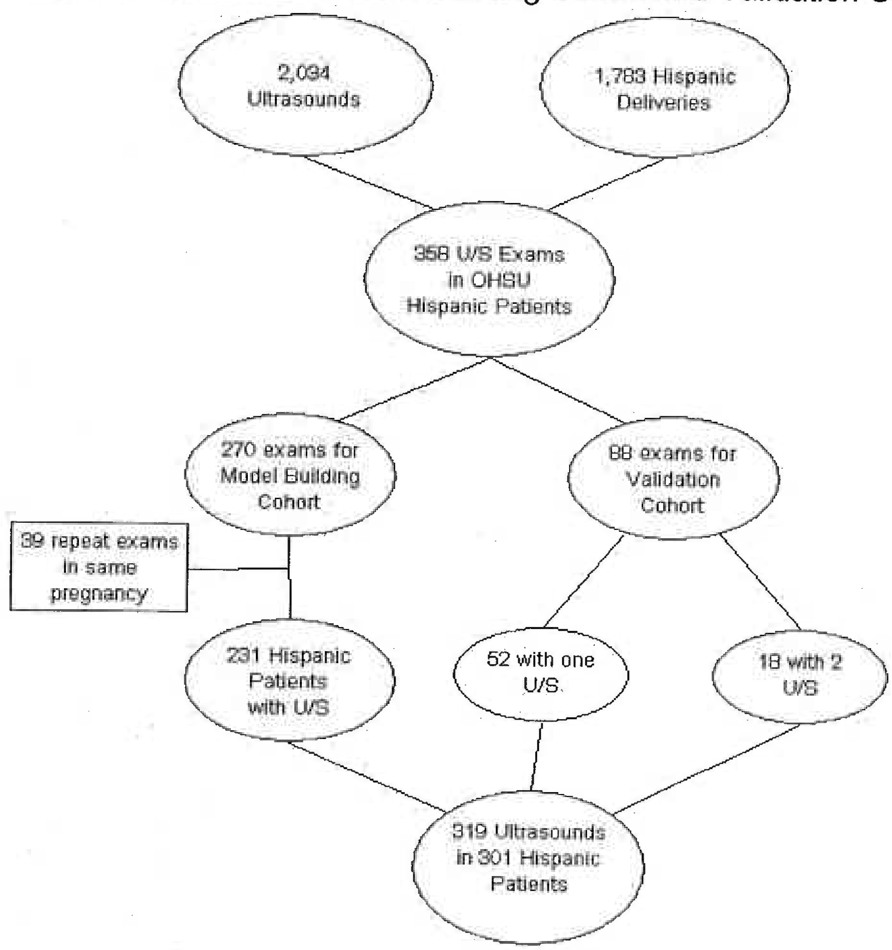
were 71 multiple Hispanic births leaving 1,783 singleton Hispanic births. The 2,034 patients from the ultrasound database (Figure 4) were then merged with the 1,783 Hispanic patients yielding 358 ultrasound examinations matched to Hispanic singleton deliveries: 270 ultrasound matches in the exploratory cohort and 88 deliveries in the validation cohort (Figure 6). Thirty-nine ultrasounds in the same pregnancy were excluded in the exploratory cohort leaving 231 ultrasound examinations from which the exploratory growth model was derived. There were 88 ultrasound examinations for the validation cohort; 52 patients had one ultrasound and 18 had two. Thus, in total, there were 319 ultrasound examinations in 301 patients.

FIGURE 5: Delivery Record Elimination Procedure



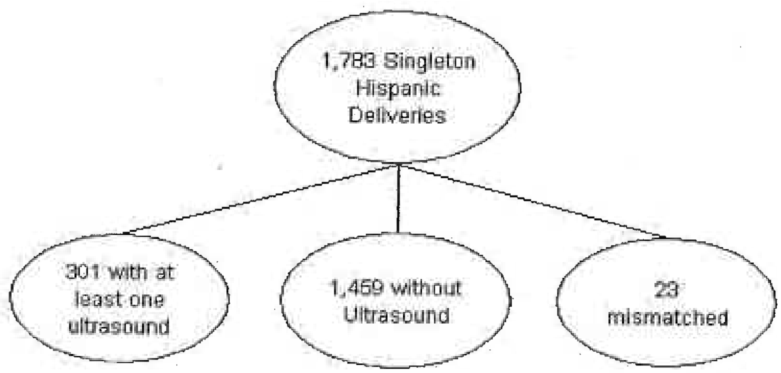
Singleton Deliveries in Hispanic Patients

FIGURE 6: Matching Process for Hispanic Deliveries to Ultrasound Examinations and Division into Model Building Cohort and Validation Cohort



In accordance with previous studies, only one ultrasound per delivering Hispanic mother was used in the generation of the OHSU Hispanic femur length curve. However, we allowed repeat examinations to remain in the validation cohort because the diagnosis of short femur length was determined for each of the 88 ultrasounds and that most closely corresponded to clinical practice. Figure 7 shows that for the 1,783 singleton Hispanic deliveries, after the exclusion of 23 mismatched patients, 301 patients had at least one ultrasound and 1,459 had no ultrasounds.

FIGURE 7: Distribution of Hispanic Deliveries for 1995-1996



Demographic Characteristics

Table 2 illustrates comparisons of demographic characteristics between the entire cohort of the Hispanic population with ultrasound at OHSU, and the entire delivering population at OHSU over the two-year time frame. Overall, the data for these characteristics were missing less than 5% due to omissions or nonsensical entries with the exception of level of education. For level of education, 22% of entries were missing for the Hispanic model building cohort and 38% were missing for the entire delivered population. A comparison of number of children born, or parity, between the populations was not performed due to presumed input or abstraction error, as there were 71 entries indicating more than 21 deliveries. Statistically significant differences between the two populations were detected for gestational age at delivery, weight gain, maternal level of education and tobacco use.

TABLE 2: Demographics for Entire Delivered Population versus Hispanic With Ultrasound Examinations

Variable	Entire Delivered n=4367 (Mean or rate) ± SD	Hispanic n=301 (Mean or Rate) ± SD	p-value
Age (Years)	25.3 ±6.1	25.9 ±6.0	0.10
Gestational age (Weeks)	38.3 ±3.2	38.7 ±2.3	0.02
Birthweight (Grams)	3,193 ±777	3,270 ±616	0.09
Weight gain (Pounds)	28.4 ±13.4	25.2 ±11.8	<0.001
Care initiated (Months)	3.2 ±1.8	3.1 ±1.7	0.66
Level of education (Years)	8.6 ±3.5	6.7 ±2.7	<0.001
Tobacco use	837/4364(19.2%)	6/301(2.0%)	<0.001
Married	2386/4367(54.6%)	175/301(58.1%)	0.24

Table 3 illustrates comparisons of demographic characteristics between the delivered Hispanic population who had an ultrasound at OHSU and those who did not have one at OHSU. Overall, the data for these variables were missing less than 5% due to omissions or nonsensical entries with the exception of level of education. For these two subsets, the only statistically significant differences were in maternal age and month care initiated.

TABLE 3: Demographics of Hispanic Population with and without Ultrasound Examinations at OHSU

Variable	Hispanic No U/S n=1459 (Mean or rate ± SD	Hispanic with U/S n=301 (Mean or rate ±SD	p-value
Age (Years)	24.1 ±5.5	25.9 ±6.0	<0.001
Gestational age (Weeks)	38.9 ±2.6	38.7 ±2.3	0.49
Birthweight (Grams)	3,309 ±641	3,270 ±616	0.34
Weight gain (Pounds)	26.1 ±12.3	25.2 ±11.8	0.28
Care initiated (Months)	3.4 ±1.8	3.1 ±1.7	0.009
Level of education (Years)	6.7 ±3.0	6.7 ±2.7	0.82
Tobacco use	32/1459(2.2%)	6/301(2.0%)	0.88
Married	853/1459(58.5%)	175/301(58.1%)	0.92

Linear Regression for the Exploratory Data Set

This exploratory data set of 231 Hispanic mothers with ultrasounds was used to generate linear, quadratic and cubic regressions of femur length versus gestational age as follows:

Linear model $FL=-1.780045+0.243016GA$

Quadratic model $FL=-3.873885+0.422568GA-0.003540GA^2$

Cubic model $FL=-5.451717+0.627312GA-0.011933GA^2+0.000109GA^3$

These models are summarized in Tables 4 and 5 and Figure 8. The regression coefficients with their standard errors and p-values for testing the coefficient equal to zero are shown in Table 4. All of the coefficients are highly significant with the exception of b_3 for the cubic model (p=0.026). We chose to use the quadratic model because of the highly significant coefficients, b_1 and b_2, and for comparison to the Hadlock growth, which is quadratic.

TABLE 4: Regression Coefficients for Linear, Quadratic and Cubic Models

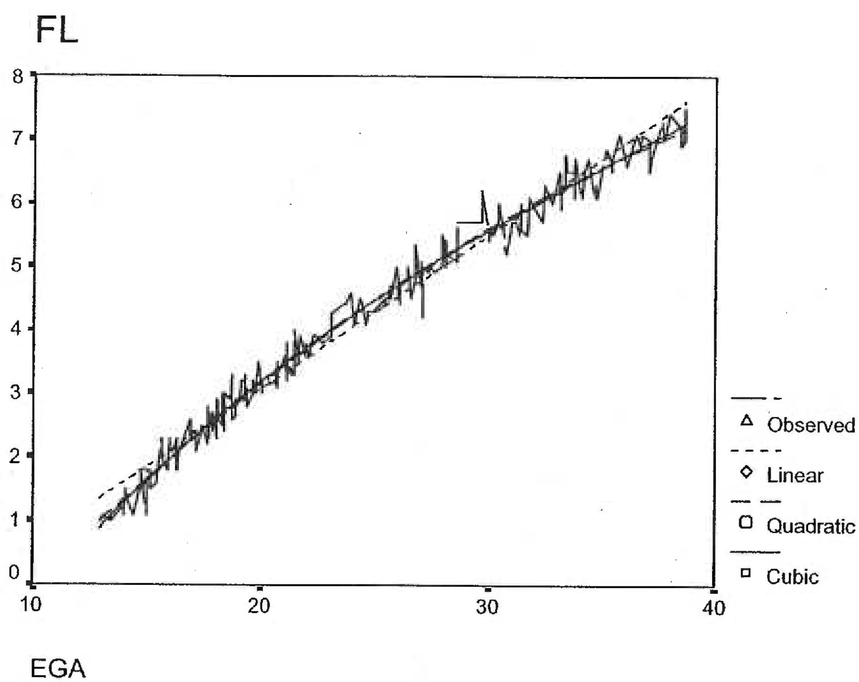
Model	b0	b1 $\pm$SE P-value[1]	b2 $\pm$SE P-value	b3 $\pm$SE P-value
Linear	-1.7800	0.2430 0.0026 0.0000		
Quadratic	-3.8739	0.4226 0.0033 0.0000	-0.0035 0.0003 0.0000	
Cubic	-5.4517	0.6273 0.0929 0.0000	-0.0119 0.0038 0.0017	0.0001 4.88E-05 0.0260

[1]P-value for test of coefficient equal to zero

TABLE 5: Comparison of Generated Models for FL on EGA in Hispanic Patients

Model	SSR	SSE	MSR	MSE	R2
Linear	683.41	18.06	683.41	0.08	0.97426
Quadratic	689.41	12.05	344.71	0.05	0.98282
Cubic	689.67	11.79	229.89	0.05	0.98319

FIGURE 8: Graphic Comparison of Linear, Quadratic and Cubic Models



A plot of the residuals for each of the three models is depicted in Figure 9. The pattern of residuals for the simple linear model does not appear to be random whereas the residual plots for both the quadratic and cubic models do appear random.

The standard FL nomogram equation model used by OHSU and other hospitals in the United States is the formula of Hadlock (1984), $FL=-3.91+0.427GA-.0034GA^2$ with standard deviation, $s_{FL.GA}$, of 0.30cm. The standard deviation for the Hispanic model is the square root of the MSE=0.0529 or 0.23cm, which is comparable to the standard deviation of the Hadlock model.

FIGURE 9: Residual Plots for Linear, Quadratic and Cubic Models

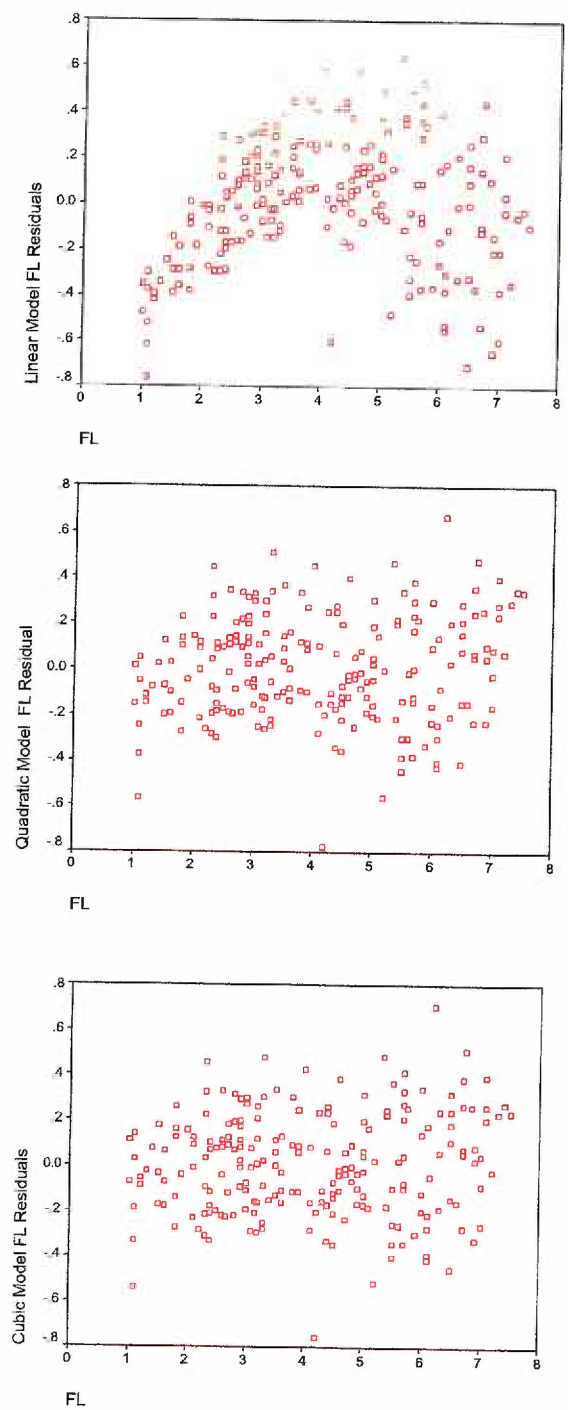


Figure 10 represents a graphical representation of the OHSU Hispanic FL normative curve with 90th percentile confidence bands. This graphical depiction is most useful for following trends in femur length growth over multiple ultrasound examinations. Figures 11 and 12 show comparisons of the means and the 5th percentile cutoffs for the Hadlock and OHSU Hispanic FL models. For a single ultrasound examination, however, the determination of whether a FL is unusually short or long for the fetal gestational age is more easily ascertained using a tabular representation of the data (Tables 6 and 7).

FIGURE 10: Graph of OHSU Hispanic FL Normative Curve with 90% Confidence Bands

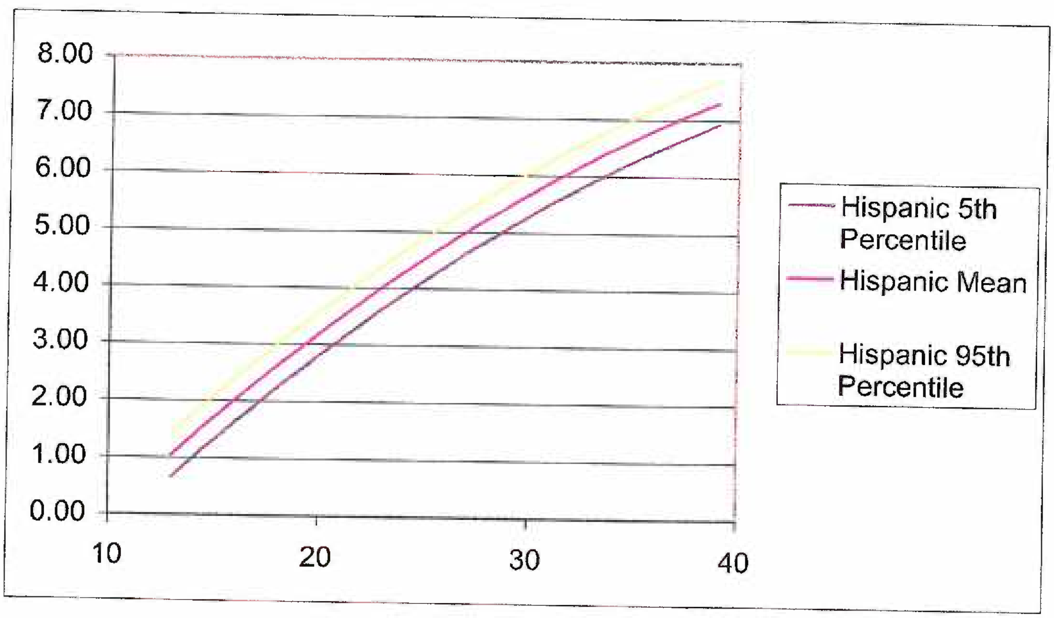


TABLE 6: Hadlock FL Regression Curve

Weeks	5th	Mean	95th
13	0.57	1.07	1.56
14	0.91	1.40	1.90
15	1.24	1.73	2.22
16	1.56	2.05	2.55
17	1.87	2.37	2.86
18	2.18	2.67	3.17
19	2.48	2.98	3.47
20	2.78	3.27	3.76
21	3.06	3.56	4.05
22	3.34	3.84	4.33
23	3.62	4.11	4.61
24	3.89	4.38	4.87
25	4.15	4.64	5.13
26	4.40	4.89	5.39
27	4.65	5.14	5.63
28	4.89	5.38	5.87
29	5.12	5.61	6.11
30	5.35	5.84	6.33
31	5.57	6.06	6.55
32	5.78	6.27	6.77
33	5.98	6.48	6.97
34	6.18	6.68	7.17
35	6.38	6.87	7.36
36	6.56	7.06	7.55
37	6.74	7.23	7.73
38	6.91	7.41	7.90
39	7.08	7.57	8.07
40	7.24	7.73	8.22

$FL=-3.91+0.427*GA-.0034*GA^2$
SD=0.30cm

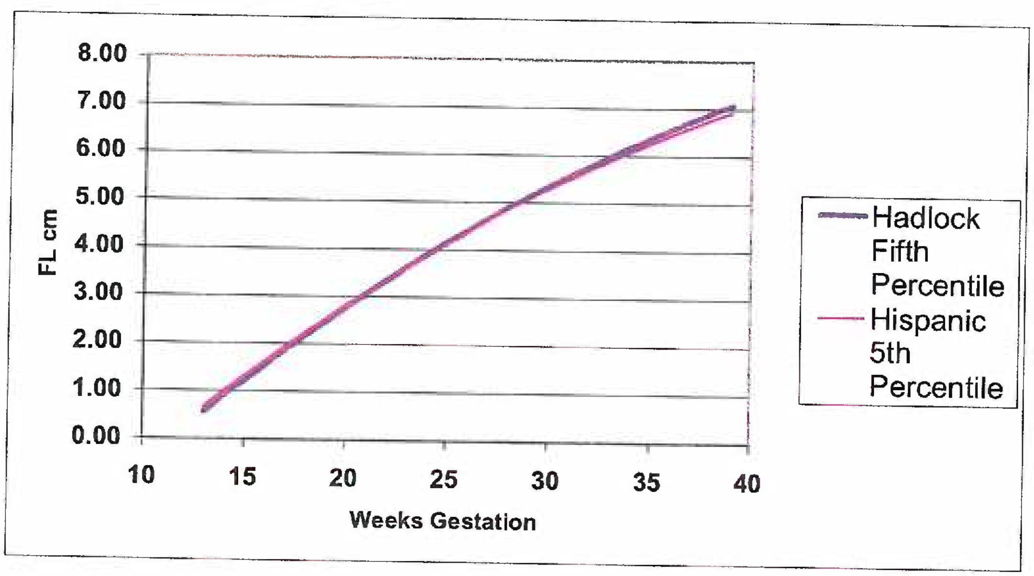
TABLE 7: OHSU Hispanic FL Regression Curve

Weeks	5th	Mean	95th
13	0.66	1.04	1.42
14	0.99	1.37	1.74
15	1.31	1.69	2.07
16	1.62	2.00	2.38
17	1.93	2.31	2.69
18	2.23	2.61	2.99
19	2.53	2.90	3.28
20	2.81	3.19	3.57
21	3.09	3.47	3.85
22	3.36	3.74	4.12
23	3.63	4.01	4.39
24	3.89	4.27	4.64
25	4.14	4.52	4.90
26	4.38	4.76	5.14
27	4.62	5.00	5.38
28	4.85	5.23	5.61
29	5.08	5.45	5.83
30	5.29	5.67	6.05
31	5.50	5.88	6.26
32	5.70	6.08	6.46
33	5.90	6.28	6.66
34	6.09	6.47	6.84
35	6.27	6.65	7.03
36	6.44	6.82	7.20
37	6.61	6.99	7.37
38	6.77	7.15	7.53
39	6.93	7.30	7.68
40	7.07	7.45	7.83

$FL=-3.87+0.423*GA-0.0035*GA^2$
SD=0.23cm

FIGURE 11: Comparison of Hadlock and OHSU Hispanic Means

FIGURE 12: Comparison of Hadlock and OHSU Hispanic 5th Percentiles



Determination of Short Femur Length for the Hispanic and Hadlock Models

The Hadlock and OHSU Hispanic regression growth models derived from the 18-month exploratory sample were used to estimate the incidence of less than 5th percentile diagnoses in the OHSU Hispanic population, as shown in Table 8. For this sample, the incidences of short femur diagnoses among Hispanic mothers undergoing an ultrasound examination were 0.095 and 0.052 for the Hadlock and OHSU Hispanic models respectively.

TABLE 8: Comparison of Models for 18-Month Hispanic Model Building Cohort

	Hadlock Normal	Hadlock Short	Totals
Hispanic Normal	208	11	219
Hispanic Short	1	11	12
Totals	209	22	231

McNemar's test, p=0.006

Table 9 represents the incidence of less than 5th percentile diagnoses in the 6-month validation cohort using the OHSU Hispanic model derived from the exploratory sample. For this validation sample the incidences of short femur diagnoses using the Hadlock and OHSU Hispanic models were 0.125 and 0.080 respectively.

TABLE 9: Comparison of Models for 6-Month Hispanic Validation Cohort

	Hadlock Normal	Hadlock Short	Totals
Hispanic Normal	77	4	81
Hispanic Short	0	7	7
Totals	77	11	88

McNemar's test, p=0.125

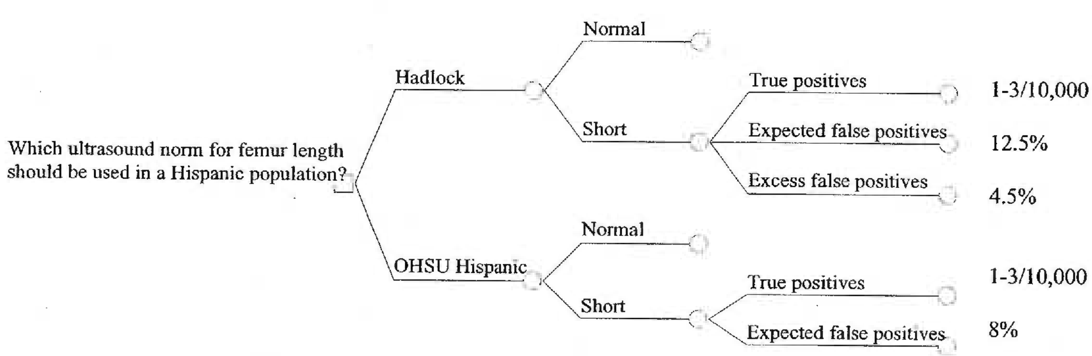
Cost Analysis

Skeletal dysplasia or dwarfism is a rare diagnosis, occurring in 1-3 per 10,000 deliveries (Romero, 1988). In the two-year time frame studied, there was only a single case of neonatal dwarfism diagnosed. This delivery occurred during the model building time frame, and was not included in the model, since the combined gestational age was not consistent with dates. Either the Hadlock or the Hispanic FL models would have detected the gross discordance in this 32-week fetus with FL of 18 weeks size.

Since dwarfism is such a rare diagnosis, it was not possible to accurately assess test performance with the data available. When ultrasound is used as a screening test for dwarfism, the vast majority of diagnoses of short femur will be false positive. Using the screen positive rates reported in Table 9, we constructed the cost analysis depicted in Figure 13. If the

Hadlock model is compared to the OHSU Hispanic model in a Hispanic population we would expect about 4.5% more screen positive diagnoses. In an institution where approximately 1000 Hispanic patients are delivered per year, if all of the patients received an ultrasound evaluation and all diagnoses of short femur resulted in a single follow up ultrasound assessment, the use of the OHSU Hispanic FL model would result in approximately 45 fewer short femur misclassifications. Using the Medicare standard reimbursement rate of $120 for each ultrasound (CPT 76805), the elimination of these misclassifications would translate into a cost savings of $5400 per one thousand patients screened.

FIGURE 13: Decision Tree for Cost Comparisons for the Two FL Nomograms



DISCUSSION

Obstetrical ultrasound may be used as a screening or a diagnostic test in pregnancy and allows assessment of fetal growth as well as anatomy. In this thesis, we address the use of ultrasound as a screening test in the assessment of fetal growth, specifically the fetal femur length. We focus on short femurs, which from a clinical standpoint, may represent a normal population variation or abnormally slow fetal growth from poor placental function or a congenital anomaly of the fetal skeletal system (dwarfism).

This study demonstrates that across the full range of gestational age, the average femur length for the fetus of a Hispanic patient at OHSU is shorter than would be predicted from the standard Hadlock reference nomogram. This results in excessive diagnoses of short femur in this ethnic population, and unnecessary costs associated with needless follow up assessment, as well as maternal anxiety over possible fetal problems.

Tables 6, 7 and 11 show that the calculated mean for each gestational week using the OHSU Hispanic is consistently 2.4-3.6% lower than that calculated using the Hadlock standard curve. However, because the conditional standard deviation for the Hispanic curve, 0.23cm, is smaller than that of the Hadlock curve, 0.30cm, and the regression coefficients for the curves are slightly different, the lower 5th percentile for the OHSU Hispanic curve is actually above that of the Hadlock curve for gestations of less than 24 weeks.

TABLE 11: Differences in Hadlock Predicted FL Values versus OHSU Hispanic Predicted FL Values

Weeks	Difference in Means %	Difference in 5th Percentile %
13	2.71	-15.05
14	2.54	-8.76
15	2.46	-5.88
16	2.42	-4.21
17	2.40	-3.11
18	2.41	-2.33
19	2.42	-1.73
20	2.45	-1.27
21	2.48	-0.88
22	2.51	-0.56
23	2.55	-0.28
24	2.59	-0.04
25	2.64	0.18
26	2.69	0.37
27	2.74	0.55
28	2.80	0.72
29	2.85	0.88
30	2.91	1.03
31	2.97	1.17
32	3.04	1.30
33	3.10	1.43
34	3.17	1.56
35	3.24	1.68
36	3.31	1.81
37	3.39	1.92
38	3.46	2.04
39	3.54	2.16
40	3.62	2.28

For the sample studied, all of the diagnoses of short FL using the Hadlock and Hispanic curves were "false positives" in the sense that none actually had dwarfism. In general diagnoses of short femur length were more commonly made after 24 weeks. When both models are applied to the validation sample, 27% of the short femur diagnoses were made prior to 24 weeks using the Hadlock model, whereas 43% of the short FL diagnoses occurred before 24 weeks gestation with the Hispanic model.

Study design issues which might have biased the results of the model created include the ultrasound elimination process, the inability to access , information regarding coexisting maternal disease, the high rates of uncertain dating among those patients undergoing ultrasound, and the low rate of ultrasound examinations in the OHSU delivering Hispanic population. The ultrasound examination elimination process, which requires consistency between the estimated gestational age by last menstrual period and composite ultrasound biometry, would tend to reduce the variation in the model population by excluding the extremes at each gestational age. This might result in confidence intervals that are narrower than that of the whole population.

The inability to access information regarding coexisting maternal disease, which might hamper fetal growth, such as diabetes or hypertension, may have resulted in calculation of population means that are actually smaller than that which would be expected in a healthy population. Although we had originally assumed that the information on coexisting maternal disease would be available from the EBC database, when we examined the actual data, we found this not to be the case. The only information regarding pregnancy complications was the rate of prematurity which, in the model building cohort was low at 7.8%, as compared to national averages of 10-12%.

A significant number of the ultrasound examinations were missing information regarding the last menstrual period (4,739/7,415 or 64%). It is not possible from the database to determine whether last menstrual period was truly unknown, or whether the sonographer performing the examination simply failed to enter the data. Since the goal was well-dated pregnancies for creation of the OHSU Hispanic curve, the exclusion of these cases seemed appropriate. We had no reason to believe that these exclusions biased the construction of the model.

In our target population, only 301 of the 1,783 (17%) Hispanic women delivering at OHSU had an ultrasound examination at OHSU prior to delivery. This number does not represent all ultrasounds obtained for Hispanic women delivering at OHSU, as it is certainly possible that many of these women received limited ultrasound examinations at their prenatal clinics to confirm dating, rather than incur the cost of a full anatomic evaluation at OHSU Radiology. Furthermore, given that 64% of the ultrasounds eligible for inclusion in the study were missing menstrual dating, it is likely that a significantly larger segment of the Hispanic population received an ultrasound evaluation. Even if the true rate were as low as 36%, the low rate of ultrasound matches to delivered Hispanic mothers would not be surprising given that the Hispanic population delivering at OHSU is generally uninsured, is composed of a significant number of migratory workers, and has significant problems with access to care and transportation.

A final study design issue that might have biased the results was the inclusion of repeat or second ultrasound studies on the same patient in the validation sample. Since an assessment of whether fetal growth is appropriate is done at every ultrasound evaluation, this seemed appropriate when designing the study. However, in retrospect, there is the possibility that the results of the second examination are correlated with the results of the first examination. To address this, the second examinations were removed from the validation cohort, leaving 70 patients with a single ultrasound examination. This resulted in a screen positive rate of 0.157 for the Hadlock model and 0.086 for the OHSU Hispanic model with a McNemar's p-value of 0.063. The screen positive rate for both models is increased, but the relative proportions of false positives remain very similar, and the inclusion of the repeat examinations did not appear to significantly bias the model performance assessment. The increase in the rate of screen positives was due to only 1 of the removed studies being a short FL and the other 16 being normals, resulting in a significant decrease in the denominator with little change in numerator.

In this study, assessment of important demographic differences between the Hispanic women with ultrasounds and the OHSU population overall revealed statistically significant differences between gestational age at delivery, weight gain, maternal level of education and maternal tobacco use. While these differences reached statistical significance, this is for the most part due to the large number of patients in the overall delivered population, rather than due to clinically significant differences in gestational age, weight gain, or maternal level of education. The Hispanic population clearly has a significantly lower rate of tobacco use than the general delivered population at OHSU for this two-year time frame. In terms of bias, the lower level of education and weight gain in the Hispanic population would tend to bias toward smaller babies, and the higher gestational age at delivery and very low level of tobacco use would tend to bias toward larger babies.

In terms of demographic differences between the Hispanic women included in the model and those without ultrasound at OHSU, the higher maternal age and earlier initiation of care would tend to bias toward larger babies in the population with ultrasound at OHSU.

Issues with Standard References

Close inspection of standard reference tables in some obstetrical ultrasound textbooks reveals some surprising findings. In a least one standard reference, it was noted that the authors were using data grouped by week of gestation for calculating weekly standard deviations, rather than the conditional deviation generated by the model. This results in very small CI at lower gestational ages, with a progressively enlarging CI with advancing gestational age, or a "megaphone-shaped" confidence interval graph. Conventional construction of the 90 percent confidence bands around the regression growth curves uses the conditional standard deviation for a given gestational age, $s_{FL.GA}$. This conditional standard deviation is assumed to be equal across the gestational ages, which was the case for our Hispanic population, but appeared not to be the case for the Hadlock population. Hadlock et al (1982) used weekly standard deviations to construct their 5th and 95th confidence limits, whereas we used the conditional standard deviation to construct the limits in Tables 6 and 7. Additionally, this same major text refers to the wrong primary reference, citing the Hadlock article of 1984, rather than the one from 1982, which actually describes the mean and standard deviation of the femur lengths for each gestational age week grouping.

It is also interesting to note that in Hadlock et al (1984) the same set of data employed to regress the dependent variable of FL on the independent variable of gestational age was used “backwards” to determine the gestational age of a fetus in utero from measured biometric parameters. While this may be a clinically useful strategy, from a statistical methodology viewpoint it is questionable since the independent variable is usually assumed to be held constant when constructing a least squares model.

SUMMARY AND CONCLUSIONS

Based on the application of the OHSU Hispanic normative regression curve to the validation sample, the addition of this normative regression equation to the ultrasound report may decrease the number of follow-up ultrasound examinations ordered. However, several issues bear consideration.

First, since the OHSU Hispanic model 5^{th} percentile cutoffs are higher than those of the Hadlock model before 24 weeks, it is possible that the use of the Hispanic model for ultrasound screening in this most common time period could actually result in more false positives.

Second, the difference between the Hadlock and Hispanic models must be taken in the context of interobserver variability. We could not calculate interobserver variability in this study, because the study was retrospective. However, previous studies have reported interobserver variability for FL measurements of 0.3mm with SD 2.16mm (Hadlock, 1982), and 0.9mm with SD 0.7mm (Smulian, 1995). Given the magnitude of the Hadlock SD relative to the mean, the Smulian estimate appears to be more precise. Comparison of the Hadlock to OHSU Hispanic cutoffs reveals that of the 28 weeks evaluated, 20 of the measurement differences in mm between the two models were less than the 0.9mm mean for interobserver variability (Tables 6 and 7).

Third, the utility of the Hispanic FL curve as a screening test for dwarfism is questionable, because dwarfism is a very rare diagnosis, found in only 1-3 per 10,000 ultrasound examinations. There was only one case in the two-year time frame of the study, and both models would have detected that case, as the diagnosis was at 32 weeks with a FL measurement of 18-week size, which is well below the 5th percentile cutoff for both models. Given the rarity of the diagnosis, it is not really possible to compare with any degree of certainty how the two models perform in terms of false negatives or missed diagnoses. Until these remaining issues can be resolved, it would be premature to recommend replacement of the current standard with this new model for the Hispanic patients receiving ultrasound at OHSU.

REFERENCES

Davis, R. O., Cutter, G. R., Goldenberg, R. L., Hoffman, H. J., Cliver, S. P., and Brumfield, C. G. 1993, "Fetal biparietal diameter, head circumference, abdominal circumference and femur length: A comparison by race and sex.": **Journal of Reproductive Medicine**, 38, p. 201-206.

Doubilet, P.M., Benson, C.B. 1993, "Improved prediction of gestational age in the third trimester." **Journal of Ultrasound in Medicine,** 12. p647-653.

Gabbe SG, Iams JD: Intrauterine growth retardation. In Iams JD, Zuspan FP (eds): Manual of Obstetrics and Gynecology. St. Louis, CV Mosby, 1990, p 169.

Guaran, R. L., Wein, P., Sheedy, M., Walstab, J., and Beischer, N. A. 1994, "Update of growth percentiles for infants born in Australian population." **Australian and New Zealand Obstetrics and Gynecology**, 34, p. 39-50.

Hadlock, F.P., Deter, R.L., Harrist, R.B., Park, S.K. 1982, "Fetal femur length as a predictor of menstrual age: Sonographically measured." **American Journal of Roentgenology**, 138, p. 875.

Hadlock, F.P., Deter, R.L., Harrist, R.B., Park, S.K. 1984, "Estimating fetal age: Computer assisted analysis of multiple fetal growth parameters." **Radiology**, 152, p. 497-501.

Helsel, D., Petitti, D. B., and Kunstadter, P. 1992, "Pregnancy among the Hmong: Birthweight, age and parity." **American Journal of Public Health**, 82, p. 1361-1364.

Hill, L. M., Guzick, D., Thomas, M. L., and Fries, J. K. 1989, "Fetal radius length: A critical evaluation of race as a factor in gestational age assessment.": **American Journal of Obstetrics and Gynecology**, 161, p. 193-199.

Hopkins, D. D. and Clarke, N. G. 1983, "Indochinese refugee fertility rates and pregnancy risk factors, Oregon.": **American Journal of Public Health**, 73, p. 1307-1309.

Jacquemyn, Y., Sys, S.U., and Verdonk, P. 2000, "Fetal biometry in different ethnic groups." **Early Human Development**, 57, p.1-13.

Jeanty, P., Rodesch, F., Delbeke, D., Dumont, J.E. 1984, "Estimation of gestational age from measurements of fetal long bones." **Journal of Ultrasound in Medicine,** 3, p.75-79.

Krieger, N. 1993, "Analyzing socioeconomic and racial/ethnic patterns in health and health care." **American Journal of Public Health**, 83, p. 1086-1087.

Lim, J.M., Hong, A.G., Raman, S. and Shyamala, N. 2000, "Relationship between fetal femur diaphysis length and neonatal crown-heel length: the effect of race." **Ultrasound in Obstetrics and Gynecology**, 15, p. 131-137.

Martin, J.A., Park, M.M., Sutton, P.D. 2002, Births: Preliminary Data for 2001. **National Vital Statistics Reports**, 50, p.1-20.

Moore, W. M. O., Bannister, R. P., Ward, B. S., Hillier, V. F., and Bamford, F. N. 1995, "Fetal and postnatal growth to age 2 years by mother's country of birth." **Early Human Development**, 42, p. 111-121.

O'Brien, J.E., Dvorin, E., Drugan, A., Yaron, Y., and Evans, M.I. 1997, "Race-ethnicity specific variation in multiple marker biochemical screening: alpha-fetoprotein, hCG, and estriol." **Obstetrics and Gynecology**, 89, p.355-358.

Otto, C. and Platt, L. D. 1991, Fetal growth and development. **Obstetrics and Gynecology Clinics of North America**, 18, p. 907-931.

Papiernik, E., Cohen, H., and Richard, A. 1986, Ethnic difference in duration of pregnancy: **Annals of Human Biology**, 13, p. 259-265.
Rimoin, D. L., Borochowitz, Z., and Horton, W. A. 1986, Short Stature - Physiology and Pathology. **Western Journal of Medicine**, 144, p. 710-721.

Romero, R., Pilu, G., Jeanty, P., Ghidini, A., Hobbins, J.C. 1988 **Prenatal Diagnosis of Congenital Anomalies.** Appleton-Lange, Norwalk, Connecticut.

Ruvolo, K.A., Filly, R.A., and Callen, P.W. 1987, Evaluation of fetal femur length for prediction of gestational age in a racially mixed obstetric population. **Journal of Ultrasound in Medicine**, 6, p. 417-419.

Shipp, T.D., Bromley, B., Mascola, M., and Benacerraf, B. 2001, Variation in fetal femur length with respect to maternal race. **Journal of Ultrasound in Medicine**, 20, p. 141-144.

Siegel, S. Castellian, N.J. 1988, **Nonparametric statistics for the behavioral sciences**. Second edition. McGraw-Hill, New York.

Scicione, A.C., Gorman, R., and Callan, N.A. 1996, Adjustment of birth weight standards for maternal and infant characteristics improves the prediction of outcome in the small for gestational age infant. **American Journal of Obstetrics and Gynecology**, 175, p.544-547.

Shiono, P. H., Klebanoff, M. A., Graubard, B. I., Berendes, H. W., and Rhoads, G. G. 1986, Birth weight among women of different ethnic groups. **Journal of the American Medical Association**, 255, p. 48-52.

Smulian, J.C., Campbell, W.A., Rodis, J.F., Feeney, L.D., Fabbri, E.L., Vintzileos, A. M. 1995. Gender-specific second-trimester biometry. **American Journal of Obstetrics and Gynecology**, 173, p.1195-1201.

Thilaganathan, B., Khare, M., and Wathen, N.C. 1998, Influence of ethnic origin on nuchal translucency screening for Down's syndrome. **Ultrasound in Obstetrics and Gynecology**, 12, p.112-114.

Ventura, S. J. and Taffel, S. M. 1985, "Childbearing characteristics of U.S.- and foreign-born Hispanic mothers.": **Public Health Reports**, 100, p. 647-652.

Yudkin P.L., Aboualfa, M., Eyre J.A., Redman, C.W.G., Wilkinson, A.R. 1987, "New birthweight and head circumference centiles for gestational ages 24 to 42 weeks." **Early Human Development**, 15, p.45-52.