OBSTETRICS

Accurate prediction of gestational age using newborn screening analyte data

Kumanan Wilson, MD, MSc, FRCPC; Steven Hawken, PhD; Beth K. Potter, PhD; Pranesh Chakraborty, MD, FRCPC, FCCMG; Mark Walker, MD, FRCS(C), MSc(epi), MScHCM; Robin Ducharme, MSc; Julian Little, MA, PhD

BACKGROUND: Identification of preterm births and accurate estimates of gestational age for newborn infants is vital to guide care. Unfortunately, in developing countries, it can be challenging to obtain estimates of gestational age. Routinely collected newborn infant screening metabolic analytes vary by gestational age and may be useful to estimate gestational age.

OBJECTIVE: We sought to develop an algorithm that could estimate gestational age at birth that is based on the analytes that are obtained from newborn infant screening.

STUDY DESIGN: We conducted a population-based cross-sectional study of all live births in the province of Ontario that included 249,700 infants who were born between April 2007 and March 2009 and who underwent newborn infant screening. We used multivariable linear and logistic regression analyses to build a model to predict gestational age using newborn infant screening metabolite measurements and readily available physical characteristics data (birthweight and sex).

RESULTS: The final model of our metabolic gestational dating algorithm had an average deviation between observed and expected gestational age of approximately 1 week, which suggests excellent predictive ability (adjusted R-square of 0.65; root mean square error, 1.06 weeks). Two-thirds of the gestational ages that were predicted by our model were accurate within ±1 week of the actual gestational age. Our logistic regression model was able to discriminate extremely well between term and increasingly premature categories of infants (c-statistic, >0.99).

CONCLUSION: Metabolic gestational dating is accurate for the prediction of gestational age and could have value in low resource settings.

Key words: algorithm, gestational age, newborn infant screening, Newborn Screening Ontario, Ontario, preterm birth

Identification of preterm birth and accurate estimates of gestational age (GA) for newborn infants is vital for several reasons.[1,2] These estimates can provide guidance as to what treatments and investigations are most appropriate for the newborn infant and can assist with accurate assessments of neurocognitive development.[3,4] Unfortunately, in developing countries, it can be challenging to obtain estimates of GA because of a lack of prenatal ultrasound dating and unreliable patient recall of menstrual period history.[5,6] Obtaining accurate estimates of GA has been recognized by the Gates Foundation as a priority for infant health. As part of their Grand Challenges Explorations 13 competition entitled "Explore New Ways to Measure Fetal and Infant Brain Development," the Foundation sought new approaches for measuring GA accurately at birth to support the creation of developmental standard curves.[7]

We postulated that a newborn infant's GA could be estimated from newborn infant analyte values in conjunction with other readily available information, such as sex and birthweight.[8,9] Analyte data are obtained from examination of dried blood spot samples taken from heel pricks typically used for newborn infant screening. Our hypothesis stemmed from our previous work that revealed a metabolic distinction between preterm children and term children, as indicated by patterns of amino acids and endocrine markers at birth.[10] We identified that metabolic patterns varied depending on the degree of prematurity. Therefore, in this study, we sought to develop an algorithm that could estimate GA at birth, based on the analytes that are obtained from newborn infant screening.

Methods

Design

We conducted a population-based cross-sectional study to predict GA with the use of newborn infant screening analyte data and readily available physical characteristics from infants who were born in the province of Ontario, Canada.

Data

We included data for infants who were born in Ontario, Canada, from April 1, 2007, to March 31, 2009, who completed newborn infant screening. Virtually all infants who are born in Ontario undergo newborn infant screening via heel prick blood spot, which is typically obtained between 24 and 72 hours of age. The Newborn Screening Ontario (NSO) program screens each infant for 29 conditions with the use of a panel of screening analytes, most of which are measured by tandem mass spectrometry. The exceptions are 17 hydroxyprogesterone (17OHP) and thyroid-stimulating hormone (TSH), which are measured using a fluorescent immunoassay (autoDELFIA, Perkin

Cite this article as: Wilson K, Hawken S, Potter B, et al. Accurate prediction of gestational age using newborn screening analyte data. Am J Obstet Gynecol 2016;214:513.e1-9.

0002-9378

http://dx.doi.org/10.1016/j.ajog.2015.10.017

TABLE 1
Measured newborn infant screening metabolites

Acyl-carnitines	C0, C2, C3, C4, C5, C6, C8, C8:1, C10, C10:1, C12, C12:1, C14, C14:1, C14:2, C16, C18, C18:1, C18:2
Amino acids	arginine, phenylalanine, alanine, leucine, ornithine, citruline, tyrosine, glycine, argininosuccinate, methionine, valine, biotinidine
Fatty acid oxidation	C3DC, C4DC, C5OH, C5DC, C6DC
Endocrine disorders	17OHP, TSH
Galactosemia and biotinidase deficiency	GALT (Galactose-1-Phosphate Uridyltransferase), biotinidase

Wilson et al. Predicting gestational age using newborn screening analyte data. Am J Obstet Gynecol 2016.

Elmer, Waltham, MA); biotinidase, measured using a colorimetric enzyme assay (Spotchek Pro; Astoria-Pacific, Inc, Clackamas, OR); and galactose-1-phosphate uridyltransferase (GALT) measured by fluorescent enzyme assay (Spotchek Pro). The analyte levels for all infants who complete screening are available in the NSO database. Broadly, the newborn infant screening analytes include acyl-carnitines, amino acids, endocrine markers, and markers of biotinidase deficiency and galactosemia (Table 1).

The NSO analyte data have been linked securely with the use of unique encoded identifiers to health administrative data at the Institute for Clinical Evaluative Sciences, which captures data on health services use, including hospitalizations, for virtually all Ontario residents. Data on birthweight, GA, ultrasound timing, and other perinatal factors were obtained from the birth admission in the Canadian Institute for Health Information's (CIHI) Discharge Abstract Database, the Ontario Health Insurance Plan database, and the newborn infant screening record. GA was based on best obstetric estimate, a combination of self-reported first day of last menstrual period and ultrasound measurement, when available. Most mothers in Ontario receive prenatal care, including ultrasound-guided gestational dating. Small for gestational age (SGA10, below 10th percentile for birthweight given gestational age) and large for gestational age (LGA90, above 90th percentile for birthweight given gestational age) were calculated based on standard cutpoints developed in a Canadian population.

Analysis

We divided our cohort of live born infants into 3 subsamples: 1 for model development, 1 to validate independently the choice of terms that were included in the final model, and 1 dataset to assess independently the performance of the final model. These subsamples were generated by randomly partitioning infants according to a 2:1:1 ratio, stratification by term, near term, premature, and extremely premature status and sex to ensure balance across the 3 subsamples.

Data preparation for regression modeling

We removed the data of infants who screened positive for any disorder from the cohort, which had the effect of removing most extreme outliers. Even after extreme outliers were removed, most analyte distributions were strongly right skewed. To pull outliers closer to the rest of the data and stabilize the variance, analyte levels were natural log transformed. We then standardized each analyte value by subtracting the sample mean (on the log scale) and dividing the result by the sample standard deviation (on the log scale), such that the resulting transformed variable had a mean of 0 and a standard deviation of 1. This allowed for easier interpretation when we compared the relative influence of analytes in a multivariable regression model, such that the regression coefficients represented the change in GA in weeks for an increase of 1 standard deviation in the (log) analyte value.

Predictive modeling

We fit a multivariable linear regression model with continuous GA in weeks as the dependent variable and used a variable selection algorithm to select terms for inclusion in the model. The full set of analyte main effects, as well as quadratic and cubic effects, was included in all models to account for a non-linear association between analyte and GA. We then conducted a backwards elimination procedure that initially included all of the main effect terms and all pairwise interactions between analytes. The Schwarz Bayesian Criterion (SBC) was used to guide the sequential removal of interaction terms from the model. SBC is a penalized likelihood criterion that quantifies how well the model fits the data, while penalizing model complexity.[11] Models with smaller SBCs are favored. Once no more interaction terms could be removed from the model based on SBC as evaluated in the model development subsample, the backwards elimination procedure was stopped. We then calculated the square root of the mean square error (RMSE) based on fitting the development models at each step of the backwards elimination in the independent validation set and choosing the model with the lowest RMSE in the validation set. The RMSE reflects how close the model estimate is to the true GA on average across all observations. Finally, the development model performance was evaluated in the test dataset, which had no role in model fitting or validation. This process provided maximum protection from overfitting and overoptimism about model performance.

Evaluation of model performance

The model built with the use of the development and validation datasets was evaluated in the test dataset in terms of adjusted *R*-square, square root-mean-square error (RMSE), and proportion of infants with predicted GA within $\pm 1, 2, 3$, and 4 weeks of true GA. RMSE is in the units of GA and hence represents the average deviation of predicted GA from actual GA over all infants in the test