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

dataset. Model performance was evaluated for all infants, for different levels of prematurity, and for infants who were small for their GA to determine whether the model performed well in babies with low birthweight/intrauterine growth restriction. We defined *prematurity* in the following manner: term, ≥37 weeks; near term, 33-36 weeks; very preterm, 28-32 weeks, and extremely preterm, <28 weeks. We also evaluated model performance according to history of maternal ultrasound during pregnancy. We categorized infants based on whether the mother received her first ultrasound within 16 weeks, 17-20 weeks, ≥21 weeks and those with no record of their mother receiving an ultrasound during pregnancy according to Ontario Health Insurance Plan claims for diagnostic ultrasound scans that were specific to pregnancy.

Model performance for classification as ≤34 or >34 weeks GA

Thirty-four weeks gestation is an important threshold because it represents the lower limit of late preterm infant period.[12,13] It is the GA after which the health risks of preterm infants are reduced, while still remaining elevated compared with term infants.[14] To classify infants according to GA ≤34 or >34 weeks, we conducted logistic regression analysis on the test data with actual GA dichotomized as ≤34 vs >34 weeks as the outcome, and the final set of predictors that was chosen for the multiple linear regression model as covariates. The logistic regression model was fit in the model development subset as mentioned earlier, then the c-statistic (area under the receiver operating characteristic curve) as well as sensitivity, specificity, positive predictive value, and proportion of infants who were classified correctly were calculated to quantify the success of the discrimination between the groups with the use of the validation subsample. The test performance was evaluated by adjustment of the GA cutpoint to determine the optimal tradeoff (higher sensitivity comes at the cost of lower specificity and lower positive-predictive value).

All analyses were conducted with SAS software (version 9.4; SAS Institute Inc, Cary, NC) and *R* (version 3.1.2).

TABLE 2
Distribution of births by sex, prematurity, and multiplicity

Variable	N (%)
Sex	
Male	128,079 (51.29)
Female	121,621 (48.71)
Prematurity categories	
Extremely preterm (≤27 wk)	555 (0.22)
Very preterm (28-32 wk)	2,616 (1.05)
Near term (33-36 wk)	16,462 (6.59)
Term (≥37 wk)	230,067 (92.14)
Small for gestational age (below 10th percentile)	
Not small for gestational age	220,167 (91.28)
Small for gestational age	21,039 (8.72)
Large for gestational age (above 90th percentile)	
Not large for gestational age	214,800 (89.05)
Large for gestational age	26,406 (10.95)
Multiple births	
No	241,206 (96.60)
Yes	8,494 (3.40)

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

This study was approved by the institutional review board at Sunnybrook Health Sciences Centre, Toronto, Canada, and by the Ottawa Health Science Network Research Ethics Board, and the Institute for Clinical Evaluative Sciences' Privacy Office.

Results

Characteristics of sample

Data were available for virtually all of the 270,000 live born infants who were delivered in Ontario between April 1, 2007, and March 31, 2009. Complete data for all newborn infant screening study analytes were available for 249,700 infants. The sample characteristics are presented in Table 2. There were 128,079 male infants (51.3%), 230,067 term infants (92.1%), 21,039 small for GA (SGA10) infants (8.7%), 26,406 large for GA (LGA90) infants (11.0%), and 8494 babies from multiple births. We randomly partitioned the dataset into 50% model development (n = 124,854), 25% validation (n = 62,412), and 25% test (n = 62,434) subsets, while maintaining the proportions of term/near term/very preterm/extremely preterm delivery and sex ratio across subsets.

Overall model performance

Our final model included 43 effects that included birthweight and sex and a total of 311 model terms, which consisted of linear, squared, and cubed main effect terms and pairwise linear interaction terms (Appendix). The 10 most predictive analytes (in terms of the change in log-likelihood) were alanine, C5, C16, C18:2, C4DC, C5DC, tyrosine, TSH, leucine and 17OHP.

Table 3 presents model performance overall and in term children (≥37 weeks) and in increasing categories of prematurity. Results are shown for the full model that considered all analytes plus sex and birthweight, for the model excluding birthweight and for a model including sex and birthweight alone.

Overall, the final model, as evaluated in the test subsample, had an adjusted *R*-square of 0.67 and a root-mean-

TABLE 3
Model performance overall and in term and preterm infants

Model	Adjusted R^2	Overall (n = 51,161)		Term (≥37 wk; n = 47,317)		Near term (33-36 wk; n = 3295)		Very preterm (28-32 wk; n = 456)		Extremely preterm (≤27 wk; n = 93)	
		Root-mean-square error, wk	Correctly classified ±1/2/3/4 wk, %	Root-mean-square error, wk	Correctly classified ±1/2/3/4 wk, %	Root-mean-square error, wk	Correctly classified ±1/2/3/4 wk, %	Root-mean-square error, wk	Correctly classified ±1/2/3/4 wk, %	Root-mean-square error, wk	Correctly classified ±1/2/3/4 wk, %
Full model	0.65	1.06	66.8/94.9/99.3/99.8	0.97	69.1/96.4/99.8/99.97	1.70	39.0/75.6/94.8/98.9	2.30	46.5/76.9/90.4/95.0	2.10	50.7/77.5/89.4/95.1
Without birthweight	0.56	1.24	61.2/91.4/98.2/99.5	1.02	64.4/94.5/99.5/99.9/	1.80	24.4/56.6/85.7/97.1/	2.60	25.3/49.2/69.7/83.7/	3.60	23.2/46.1/61.5/73.6/
Sex and birthweight only	0.54	1.26	58.2/90.73/98.1/99.5/	1.11	61.3/94.1/99.6/99.9/	2.30	21.0/50.1/81.1/99.6/	3.00	24.0/50.3/50.1/73.3/	1.90	44.4/78.2/92.3/97.9/

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

square error (RMSE) of 1.06 (meaning the average deviation between observed and expected GA was approximately 1 week), with two-thirds of predicted GAs falling within ±1 week of actual GA (Table 2). In term children, 69% of infant GAs were predicted within ±1 week, and 96% were predicted within ±2 weeks. In near term infants, 39% were predicted within ±1 week, and 76% were predicted within ±2 weeks. In very preterm infants, 51% were predicted within ±1 week, and 77% were predicted within ±2 weeks.

Model performance in subgroups

The overall RMSE in low birthweight infants (SGA10) was 1.34, compared with 1.03 in non-SGA10 infants across all categories of prematurity. However, the increased prediction error was limited to term children (≥37 weeks), because the model performed slightly better in every category of SGA10 infants who were preterm (<37 weeks).

Table 4 provides a breakdown of the estimated category of GA compared with the actual category of GA. GA for term SGA10 infants tended to be underestimated by the model, which resulted in some SGA10 infants (10%) being misclassified as near term. However, <0.1% were misclassified as very preterm, and none were misclassified as extremely preterm (Table 5). Conversely, the model tended to overestimate GA in infants classified as LGA90. For example, >80% of LGA90 near term babies were misclassified as full term.

For comparison, a model that included only sex and birthweight had an RMSE of 1.26, and a model that included sex and all of the analytes (but not birthweight) had an RMSE of 1.23, compared with an RMSE of 1.05 for the full model that included sex, birthweight, and analytes.

Model performance for classification as ≤34 or >34 weeks GA

In the test data, the overall c-statistic (area under the ROC curve; Figure) was 0.991, which suggests excellent discrimination of GA of ≤34 vs >34 weeks. The test performance was evaluated by adjustment of the predicted probability

cutpoint of the logistic model to determine the optimal tradeoff between sensitivity and specificity. For example, the performance of the model in discriminating between ≤34 vs >34 weeks had specificity of 99.5%, positive-predictive value of 80.9%, and 98.9% of all infants were correctly classified when sensitivity was 80% (ie, 80% of infants with GA ≤34 weeks were correctly identified by the model). Table 6 presents specificity, positive-predictive value, and percentage correctly classified for benchmark sensitivities of 50-95%.

TABLE 4
Agreement of actual gestational age category and predicted gestational age category

	Predicted, %				
Actual gestational age, wk	≤27	28-32	33-36	≥37	Total
≤27	79.3	20.0	0.0	0.7	100
28-32	8.1	66.7	21.9	3.3	100
33-36	0.0	3.6	59.7	36.7	100
≥37	0.0	0.0	2.0	98.0	100

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

Model performance based on timing of dating ultrasound scan

In the full analysis cohort, 98.7% had at least 1 ultrasound scan; 69.4% had an ultrasound scan performed in the first 16 weeks of gestation; 83.5% had an ultrasound scan in the first 18 weeks of gestation, and 92.7% had an ultrasound scan in the first 20 weeks of gestation. In the model testing subset, the RMSE was 1.06 for those who had ultrasound scans in the first 16 weeks of gestation; 1.01 for weeks 17-20, and 1.11 for ≥21 weeks. If there was no code for ultrasound scan, the RMSE was 1.13.

Comment

In this study, we demonstrated the potential value of analytes that were derived from blood spots typically used for newborn infant screening to predict GA in the newborn infant. The model we developed, which used these analytes in combination with sex and birthweight, is able to predict continuous GA within about ±1 week overall and within ±1 to 2 weeks in near term and very preterm babies. The model showed excellent discrimination for classification of infants as >34 vs ≤34 weeks.

There is a potentially substantial value to the use of the blood spot–derived analytes for the estimation of GA. Although the current standard method for the determination of GA, first-trimester ultrasound scanning,[15-19] requires interpretation by a specialized physician and requires equipment that may not be available readily in resource-poor settings, analyses based on blood spots could be automated fully and standardized for this application. Other methods of the establishment of GA also have limitations.[20] Reliable records of last menstrual period may not be available in settings in which there is no prenatal care. Even when last menstrual period data are available, it may not provide an accurate estimate of GA.[21] Assessment of anterior lens capsule vascularity has been used as an alternative mechanism for postnatal GA dating. However, this approach is difficult in preterm children. A combination of physical and neurologic assessments, such as the New Ballard Score and the Dubowitz GA assessment, have emerged as the standard for postnatal GA dating.[17,22] However, these may be difficult for nonpediatricians to perform and have suboptimal interrater reliability scores.[23-26] They are not as accurate as prenatal ultrasound scanning,[27-29] have limitations at the extremes of GA, in critically ill infants, and accuracy may vary by ethnicity.[30-32] The main limitation to the use of blood spots is the availability of tandem mass spectrometers or other necessary devices. There have been advances in the development of portable tandem mass spectrometer devices that may offer the opportunity to better operationalize metabolic gestational dating in practice. In the absence of these, blood spot cards could be shipped to a setting where the necessary analytic machinery is available.

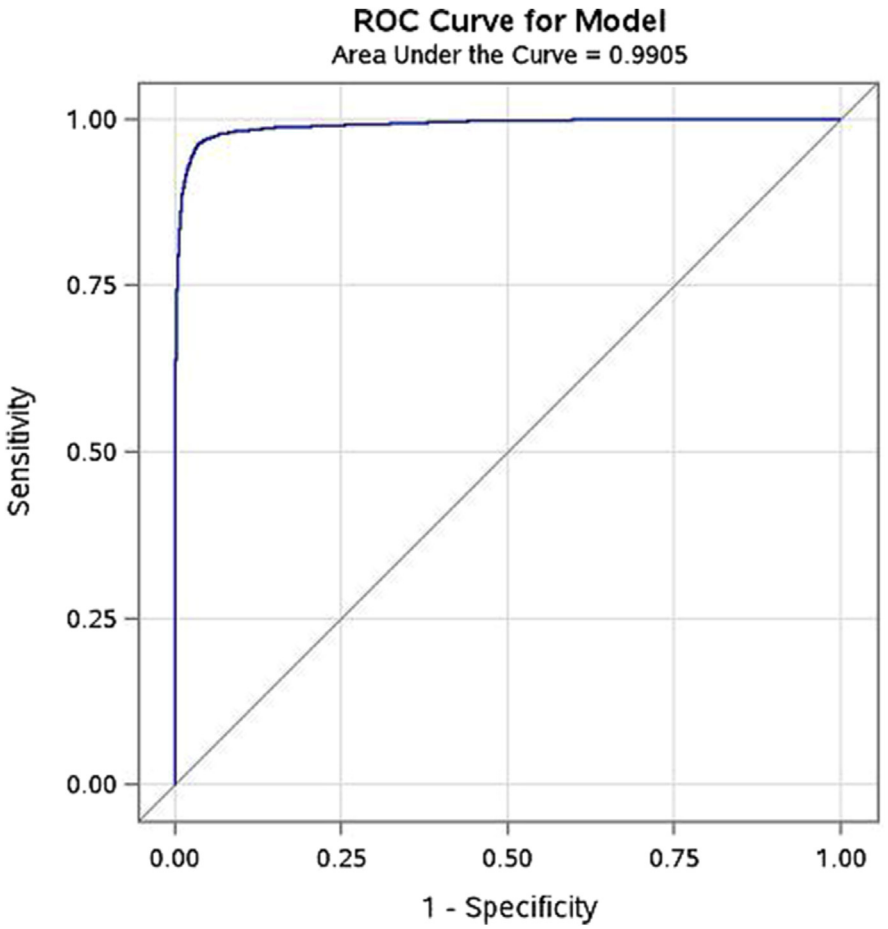
In our previous work, we identified variation in analyte levels (amino acids, endocrine markers, enzymes) based on degree of preterm birth and demonstrated heat map differences (correlations between analytes) based on categories of preterm birth.[10] We hypothesized that the differences in metabolic profile could be due to either lack of maturation of organs/pathways (eg, TSH lower in preterm children) or catabolic stress in preterm children (resulting in, for example, elevation in 17OHP).[33-36] However, low birthweight, term children are also at risk of

TABLE 5
Agreement of actual gestational age category and predicted gestational age category for small-for-gestational-age (below 10th percentile) infants

	Predicted, %				
Actual gestational age, wk	≤27	28-32	33-36	≥37	Total
≤27	100.0	0.0	0.0	0.0	100
28-32	22.7	75.0	2.3	0.0	100
33-36	0.0	14.6	79.9	5.5	100
≥37	0.0	0.1	10.4	89.5	100

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

FIGURE
Receiver operating characteristic curve for full model



The receiver operating characteristic curve represents the trade-off between false-positive and true-positive rates over all possible cutoffs of predicted probability from the logistic model. The *diagonal straight line* represents random chance. The higher the lift of the receiver operating characteristic curve from the diagonal, the better the discrimination of the model. This is represented by the area under the receiver operating characteristic curve, which is equivalent to the c-statistic for the logistic regression model.

ROC, receiver operating characteristic.

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

TABLE 6
Sensitivity, specificity, and positive-predictive value for the classification of infants as gestational age >34 vs ≤34 weeks

Sensitivity, %	Specificity, %	Positive-predictive value, %	Correctly classified, %
50	99.9	96.9	98.5
60	99.9	94.3	98.8
70	99.8	89.5	98.9
80	99.5	80.9	98.9
90	98.6	65.8	98.4
95	97.1	48.8	97.0

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

experiencing catabolic stress, and it is important to be able to distinguish these children from preterm births. Our model appears to distinguish these children effectively. Analytes plus sex had a higher predictive value than sex and birthweight alone in all children and in SGA10 term children. The addition of analytes into a model with sex and birthweight sharply improved the predictive value of the model. Perhaps most importantly, in term SGA10 children (who are likely to be at risk of catabolic stress and potentially misclassified as preterm), the model accurately identified approximately 90% of them as being term. This strongly suggests that factors other than catabolic stress are responsible for the different analyte patterns in preterm children.

Strengths of our analyses are that the large sample size and computing power enabled us to partition our data and to use a sound variable selection, internal validation, and test performance strategy to avoid potential overfitting. With >30 candidate analytes to evaluate, interactions among analytes and nonlinear relationships quickly result in a vast number of variables to consider in regression modeling. We were able to balance the need for an accurate model, to manage hundreds of candidate variables (while avoiding overfitting the model to the data), and to end up with useful model with reproducible performance characteristics. Our gold standard assessment of GA was based on best obstetric estimate. Because approximately 70% of pregnancies in Ontario have at least 2 prenatal ultrasound scans and 99.4% have at least 1 scan, the vast majority of the GA estimates likely would be informed by ultrasound scans.[37] When examining billing data on dating ultrasound scans in our cohort, we found that 93% of the patients had ultrasound scans within the first 20 weeks and that the model performed better on those patients with ultrasound scans than on those who did not have them.

A potential limitation of our analysis is the possibility that covariates at the infant, maternal, and blood spot sample level could impact the estimate of GA.[38]

In resource-poor settings, the effect of concomitant illness on analyte profiles, HIV in particular, would need to be accounted for.[39] Our models appear to predict GA less accurately in increasingly preterm children, which may be due to a combination of the smaller sample size for preterm infants and also that these infants may have more variable newborn infant screening analyte levels because of factors such as the infant's physiology, feeding status, and timing of sample collection.

Future studies should examine the impact of important infant, maternal, birth, and sample covariates on the predictive model. The impact of other variables that were collected in expanded newborn infant screening programs should also be assessed. Our model should be validated in other international settings in which newborn infant screening is being conducted.[40-43] Ultimately, a valid model should be tested in low-resource settings for which biobank cord blood and/or heel prick blood spot samples and dating ultrasound scans are available in a sample population.

If a globally valid algorithm can be developed, we envision that the following scenario could be realized: An infant is born in a resource-poor setting. Ideally, a blood spot sample is obtained immediately after birth from a heel prick. Samples potentially could also be obtained from heel pricks after birth or from cord blood. The blood spot sample is analyzed by a portable device or shipped to a center where the necessary equipment is available. Analyte values from this analysis are combined with, when available, data entered by a health care provider. This will permit modification of the algorithm so that the GA estimate is tailored to be as accurate as possible for that specific infant. Accurate information on GA for an infant will then guide care providers to the most appropriate treatments and assessments for the infant's category of prematurity. There are many important obstacles to the achievement of this objective, which include the cost of testing (NSO costs are $55 Canadian per child for the analytes included in the model), the fact that many infants in resource-poor countries are discharged at <24 hours, NSO analytes typically are obtained 24-72 hours after birth, and issues around standardization of tests. The merits of this technology, both accuracy and feasibility, should be compared with existing strategies for the estimation of GA. ■

References

1. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371:75-84.
2. Muglia LJ, Katz M. The enigma of spontaneous preterm birth. N Engl J Med 2010;362:529-35.
3. Barros FC, Bhutta ZA, Batra M, Hansen TN, Victora CG, Rubens CE. Global report on preterm birth and stillbirth (3 of 7): evidence for effectiveness of interventions. BMC Pregnancy Childbirth 2010;10(suppl 1):S3.
4. Palmer PG, Dubowitz LM, Verghote M, Dubowitz V. Neurological and neurobehavioural differences between preterm infants at term and full-term newborn infants. Neuropediatrics 1982;13:183-9.
5. Lawn JE, Gravett MG, Nunes TM, Rubens CE, Stanton C. Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. BMC Pregnancy Childbirth 2010;10(suppl 1):S1.
6. Rubens CE, Gravett MG, Victora CG, Nunes TM. Global report on preterm birth and stillbirth (7 of 7): mobilizing resources to accelerate innovative solutions (Global Action Agenda). BMC Pregnancy Childbirth 2010;10(suppl 1):S7.
7. Explore New Ways to Measure Fetal and Infant Brain Development (Round 13). Grand Challenges Exploration. Bill & Melinda Gates Foundation. (http://gcgh.grandchallenges.org/challenge/explore-new-ways-measure-fetal-and-infant-brain-development-round-13) Accessed December 2, 2015.
8. Oladipo OO, Weindel AL, Saunders AN, Dietzen DJ. Impact of premature birth and critical illness on neonatal range of plasma amino acid concentrations determined by LC-MS/MS. Mol Genet Metab 2011;104:476-9.
9. Slaughter JL, Meinzen-Derr J, Rose SR, et al. The effects of gestational age and birth weight on false-positive newborn-screening rates. Pediatrics 2010;126:910-6.
10. Wilson K, Hawken S, Ducharme R, Potter BK, Little J, Thebaud B, Chakraborty P. Metabolomics of prematurity: Analysis of patterns of amino acids, enzymes and endocrine markers by categories of gestational age. Pediatr Res 2014;75:367-73.
11. Schwarz G. Estimating the dimension of a model. Ann Statist 1978;6:461-4.
12. Kugelman A, Colin AA. Late preterm infants: near term but still in a critical developmental time period. Pediatrics 2013;132:741-51.
13. Bakewell-Sachs S. Near-term/late preterm infants. Newborn Infant Nurs Rev 2007;7:67-71.
14. Whyte R. Safe discharge of the late preterm infant. Paediatr Child Health 2010;15:655-66.
15. Saltvedt S, Almström H, Kublickas M, Reilly M, Valentin L, Grunewald C. Ultrasound dating at 12−14 or 15−20 weeks of gestation? A prospective cross-validation of established dating formulae in a population of in-vitro fertilized pregnancies randomized to early or late dating scan. Ultrasound Obstet Gynecol 2004;24:42-50.
16. Kalish RB, Thaler HT, Chasen ST, et al. First- and second-trimester ultrasound assessment of gestational age. Am J Obstet Gynecol 2004;191:975-8.
17. Ballard J, Khoury J, Wedig K, Wang L, Eilers-Walsman B, Lipp R. New Ballard Score, expanded to include extremely premature infants. J Pediatr 1991;119:417-23.
18. Rossavik IK, Fishburne JI. Conceptional age, menstrual age, and ultrasound age: a second-trimester comparison of pregnancies of known conception date with pregnancies dated from the last menstrual period. Obstet Gynecol 1989;73:243-9.
19. Persson PH, Weldner BM. Reliability of ultrasound fetometry in estimating gestational age in the second trimester. Acta Obstet Gynecol Scand 1986;65:481-3.
20. Kramer MS, Papageorghiou A, Culhane J, et al. Challenges in defining and classifying the preterm birth syndrome. Am J Obstet Gynecol 2012;206:108-12.
21. Kramer MS, McLean FH, Boyd ME, Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. JAMA 1988;260:3306-8.
22. Dubowitz LM, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. J Pediatr 1970;77:1-10.
23. Dubowitz LM, Dubowitz V, Palmer P, Verghote M. A new approach to the neurological assessment of the preterm and full-term newborn infant. Brain Dev 1980;2:3-14.
24. Dubowitz LM, Dubowitz V, Goldberg C. A comparison of neurological function in growth-retarded and appropriate-sized full-term newborn infants in two ethnic groups. S Afr Med J 1982;61:1003-7.
25. Dubowitz L, Ricciw D, Mercuri E. The Dubowitz neurological examination of the full-term newborn. Ment Retard Dev Disabil Res Rev 2005;11:52-60.
26. Alexander GR, de Caunes F, Hulsey TC, Tompkins ME, Allen M. Validity of postnatal assessments of gestational age: a comparison of the method of Ballard et al and early ultrasonography. Am J Obstet Gynecol 1992;166:891-5.
27. Alexander GR, Allen MC. Conceptualization, measurement, and use of gestational age: I, clinical and public health practice. J Perinatal 1995;16:53-9.
28. Allen MC. Assessment of gestational age and neuromaturation. Ment Retard Dev Disabil Res Rev 2005;11:21-33.

29. Wariyar U, Tin W, Hey E. Gestational assessment assessed. Arch Dis Child Fetal Neonatal Ed 1997;77:F216-20.
30. Alexander S, Buekens P, Blondel B, Kaminski M. Is routine antenatal booking vaginal examination necessary for reasons other than cervical cytology if ultrasound examination is planned? BJOG 1990;97:365-6.
31. Sanders M, Allen M, Alexander GR, et al. Gestational age assessment in preterm neonates weighing less than 1500 grams. Pediatrics 1991;88:542-6.
32. Butler AS, Behrman RE. Preterm birth: causes, consequences, and prevention. Washington (DC): National Academies Press; 2007.
33. Fisher DA. The hypothyroxinemia [corrected] of prematurity. J Clin Endocrinol Metab 1997;82:1701-3.
34. Linder N, Davidovitch N, Kogan A, et al. Longitudinal measurements of 17alpha-hydroxyprogesterone in premature infants during the first three months of life. Arch Dis Child Fetal Neonatal Ed 1999;81:F175-8.
35. Ersch J, Beinder E, Stallmach T, Bucher HU, Torresani T. 17-Hydroxyprogesterone in premature infants as a marker of intrauterine stress. J Perinat Med 2008;36:157-60.
36. Scott SM, Cimino DF. Evidence for developmental hypopituitarism in ill preterm infants. J Perinatol 2004;24:429-34.
37. Guttmann A, Vermeulen M, Simeonov D, Walker M, Ray J. Are rates of prenatal ultrasound a valid measure of health system overuse? Institute for Clinical Evaluative Science Policy Brief. Available at: http://www.ices.on.ca/~/media/Files/Briefing-Notes/ICES ECFA Policy Brief Prenatal Ultrasound 2012-04-04.ashx. Accessed: July 2, 2015.
38. Ryckman KK, Berberich SL, Shchelochkov OA, Cook DE, Murray JC. Clinical and environmental influences on metabolic biomarkers collected for newborn screening. Clin Biochem 2013;46:133-8.
39. Kirmse B, Hobbs CV, Peter I, et al. Abnormal newborn screens and acylcarnitines in HIV exposed. Pediatr Infect Dis J 2013;32:146-50.
40. Therrell BL, Adams J. Newborn screening in North America. J Inherit Metab Dis 2007;30:447-65.
41. Padilla CD, Therrell BL. Newborn screening in the Asia Pacific region. J Inherit Metab Dis 2007;30:490-506.
42. Bodamer OA, Hoffmann GF, Lindner M. Expanded newborn screening in Europe 2007. J Inherit Metab Dis 2007;30:439-44.
43. Borrajo GJ. Newborn screening in Latin America at the beginning of the 21st century. J Inherit Metab Dis 2007;30:466-81.

Author and article information

From the Clinical Epidemiology Program, Ottawa Hospital Research Institute (Drs Wilson, Hawken, Walker, and Ms Ducharme); the Institute for Clinical Evaluative Sciences (Drs Wilson, Hawken, Potter, and Ms Ducharme); the School of Epidemiology, Public Health and Preventive Medicine (Drs Wilson, Hawken, Potter, and Little) and the Department of Medicine (Dr. Wilson), the Department of Pediatrics (Dr. Chakraborty) and the Department of Obstetrics & Gynecology (Dr. Walker), University of Ottawa; Children's Hospital of Eastern Ontario Research Institute (Drs Wilson, Hawken, Chakraborty, and Walker); and Newborn Screening Ontario (Drs Potter and Chakraborty), Ottawa, Ontario, Canada.

Received Aug. 14, 2015; revised Oct. 14, 2015; accepted Oct. 18, 2015.

Supported by This study was funded by a Grand Challenges Exploration grant from the Bill & Melinda Gates Foundation (OPP1119206) and by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC).

The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. Parts of this material are based on data and information compiled and provided by the Canadian Institute for Health Information (CIHI). However, the analyses, conclusions, opinions and statements expressed herein are those of the author and not necessarily those of CIHI.

The authors report no conflict of interest.

Corresponding author: Kumanan Wilson, MD, MSc, FRCPC. kwilson@ohri.ca

APPENDIX
Predictors in the full model

Categorical: SEX

Linear, quadratic and cubic(x, x^2 and x^3 included for each covariate):

BIRTHWEIGHT ALA ARG BIO C0 C2 C3 C4 C4OH C5 C6 C8 C8:1 C10 C10:1 C12 C12:1 C14 C14:1 C14:2 C16 C18 C18:1 C18:2 C3DC C4DC C5OH C5DC C6DC CIT GLY LEU MET ORN PHE GALT TSH TYR VAL 17OHP C16OH C16:1OH C18OH C18:1OH C5:1

Interactions:

BIRTHWEIGHT*SEX
BIRTHWEIGHT*ALA
BIRTHWEIGHT*ARG
ARG*BIO
BIO*C0
BIRTHWEIGHT*C2
C2*SEX
ALA*C2
ARG*C2
BIRTHWEIGHT*C3
C3*SEX
BIO*C3
C0*C3
C2*C3
C2*C4OH
C3*C4OH
BIRTHWEIGHT*C5
BIRTHWEIGHT*C6
C0*C6
C2*C6
C2*C8
ALA*C8:1
C0*C8:1
BIRTHWEIGHT*C10
ALA*C10
C8*C12
C12:1*SEX
C4OH*C12:1
BIRTHWEIGHT*C14
BIRTHWEIGHT*C14:1
C3*C14:1
C8:1*C14:1
BIRTHWEIGHT*C14:2
C2*C14:2
C16*SEX
ALA*C16
BIO*C16
C2*C16
C6*C16
C14:2*C16
C2*C18
C12*C18:1
C18:2*SEX
ARG*C18:2
C3*C18:2
LEU*PHE
MET*PHE
C8:1*C18:2
C12:1*C18:2
C16*C18:2
C18*C18:2
BIRTHWEIGHT*C3DC
C8*C3DC
C8:1*C3DC
C12*C3DC
C18:2*C3DC
C2*C4DC
C5*C4DC
C12:1*C4DC
C14*C4DC
C16*C4DC
C18:1*C4DC
C4OH*C5OH
C14:1*C5OH
C18:2*C5OH
BIRTHWEIGHT*C5DC
ARG*C5DC
BIO*C5DC
C12*C5DC
C18:1*C5DC
ALA*C6DC
C0*C6DC
C2*C6DC
C4OH*C6DC
C8:1*C6DC
C14:1*C6DC
C16*C6DC
C18:1*C6DC
C3DC*C6DC
C2*CIT C5*CIT
C3DC*CIT
C4DC*CIT
BIRTHWEIGHT*GLY
C0*GLY
C2*GLY
C3*GLY
C16*GLY
C18:2*GLY
C6DC*GLY
CIT*GLY
BIRTHWEIGHT*LEU
C2*LEU
C3*LEU
C4DC*LEU
C6DC*LEU
C3*MET
C10*MET
C12*MET
C18:1*MET
BIRTHWEIGHT*ORN
BIO*ORN
C0*ORN
C2*ORN
C3*ORN
C5*ORN
C8*ORN
C12*ORN
C14:1*ORN
C18*ORN
C18:1*ORN
C4DC*ORN
C5DC*ORN
CIT*ORN
GLY*ORN
PHE*SEX
ALA*PHE
BIO*PHE
C18*PHE
C4DC*PHE
C6DC*PHE
GLY*PHE
BIRTHWEIGHT*GALT
C14:2*GALT
C16*GALT
BIRTHWEIGHT*TSH
C6*TSH
C18:2*TSH
C4DC*TSH
C5DC*TSH
CIT*TSH
GLY*TSH
ORN*TSH
GALT*TSH
BIRTHWEIGHT*TYR
ALA*TYR
C2*TYR
C6*TYR
C12:1*TYR
C4DC*TYR
C6DC*TYR
CIT*TYR
MET*TYR
ORN*TYR
GALT*TYR
TSH*TYR
BIRTHWEIGHT*VAL
BIO*VAL
C2*VAL
C5*VAL
C8:1*VAL
C14:1*VAL
C18:2*VAL
C5DC*VAL
LEU*VAL
MET*VAL
TYR*VAL
BIRTHWEIGHT*17OHP
C2*17OHP
C4OH*17OHP
C8:1*17OHP
C12:1*17OHP
C4DC*17OHP
C6DC*17OHP
CIT*17OHP
LEU*17OHP
MET*17OHP
TYR*17OHP
VAL*17OHP
C2*C16:1OH
GLY*C16:1OH
BIRTHWEIGHT*C5:1
C2*C5:1
C5DC*C5:1

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