Validation of gestational age determination from ultrasound or a metabolic gestational age algorithm using exact date of conception in a cohort of newborns conceived using assisted reproduction technologies

Steven Hawken, PhD; Brieanne Olibris, PhD; Robin Ducharme, MSc; A. Brianne Bota, PhD; Jeffrey C. Murray, MD; Beth K. Potter, PhD; Mark Walker, MD; Pranesh Chakraborty, MD; Kumanan Wilson, MD

BACKGROUND: Accurate estimates of gestational age in pregnancy are important for the provision of optimal care. Although current guidelines generally recommend estimating gestational age via first-trimester ultrasound measurement of crown–rump length, error associated with this method can range from 3 to 8 days of gestation. In pregnancies resulting from assisted reproductive technology, estimated due date can be calculated on the basis of the age of the embryo and the date of embryo transfer, arguably providing the most accurate estimates possible. We have developed and extensively validated statistical models to estimate gestational age postnatally using metabolomic markers from blood samples in combination with clinical and demographic data. These models have shown high accuracy compared with first-trimester ultrasound, the recommended method for estimating gestational age in spontaneous pregnancies. We hypothesized that gestational age derived from date and stage of embryo at transfer in newborns conceived using assisted reproduction therapy would provide the most accurate reference standard possible to evaluate and compare the accuracy of both first-trimester ultrasound and metabolic model—based gestational dating.

OBJECTIVE: This study aimed to validate both first-trimester ultrasound dating and postnatal metabolic gestational age estimation models against gestational age derived from date and stage of embryo at transfer in a cohort of newborns conceived via assisted reproductive technology, both overall and in important subgroups of interest (preterm birth, small for gestational age, and multiple birth).

STUDY DESIGN: This was a retrospective cohort study of infants born in Ontario, Canada between 2015 and 2017 and captured in the provincial birth registry. Spontaneous conceptions were randomly partitioned into a model derivation sample (80%) and a test sample (20%) for model validation. A cohort of assisted conceptions resulting from fresh embryo transfers was derived to evaluate the accuracy of both ultrasound and model-based gestational dating. Postnatal gestational age estimation models were developed with multivariable linear regression using elastic-net regularization. Gestational age estimates from dating ultrasound and from postnatal metabolomic models were compared with date of embryo transfer reference gestational age in the independent test cohorts. Accuracy was quantified by calculating mean absolute error and the square root of mean squared error.

RESULTS: Our model derivation cohort included 202,300 spontaneous conceptions, and the testing cohorts included 50,735 spontaneous conceptions and 1924 assisted conceptions. In the assisted conception cohort, first-trimester dating ultrasound was accurate to within approximately ±1.5 days compared with date of embryo transfer reference overall (mean absolute error, 0.21 [95% confidence interval, 0.20–0.23]). When compared with gestational age derived from date of embryo transfer, the metabolomic estimation models were accurate to within approximately ±5 days overall (0.79 [0.76–0.81] weeks). When ultrasound was used as the reference in validating the metabolomic model, the mean absolute error was slightly higher overall (0.81 [0.78–0.84] weeks). In general, the accuracy of gestational age estimates derived from ultrasound or metabolomic models was highest in term infants and lower in preterm and small-for-gestational-age newborns.
CONCLUSION: Our findings support the accuracy of ultrasound as a gestational age dating tool. They also support the potential utility of metabolic gestational age dating algorithms in settings where ultrasound or other accurate methods of estimating gestational age are not available because of lack of infrastructure or specialized training (eg, low-income countries). However, the accuracy of metabolic model—based dating was generally lower than that of ultrasound.

Key words: assisted reproduction, epidemiology, gestational age, in vitro fertilization, prediction modeling, ultrasonography

The provision and planning of appropriate obstetrical and neonatal care, timing and interpretation of antepartum tests, assessments of fetal growth and development, and timely provision of interventions related to preterm and postterm birth are all dependent on valid estimates of due date and gestational age (GA). Current guidelines recommend the use of first-trimester ultrasound to measure crown–rump length when possible to obtain the best estimate of GA, but if the pregnancy was conceived through in vitro fertilization, date of embryo transfer should be used. However, there is still error associated with first-trimester ultrasound dating, which has been reported to be accurate to within approximately 3 to 8 days. Furthermore, differences have been shown in the accuracy and consistency of fetal biometry measurements between institutions and providers, with variation in estimates of GA of up to 10 days. Accuracy is strongly dependent on the GA at which the fetal biometry measurement was taken.

Assisted reproductive technologies (ARTs) are a collection of treatments available to assist individuals in conceiving. The American College of Obstetricians and Gynecologists recommends that estimated due date in pregnancies resulting from ART be calculated on the basis of the age of the embryo and the date of embryo transfer (DET), accounting for the days before ovulation in the involved menstrual cycle (oocyte retrieval+14 days). This is arguably the most accurate method possible for determining GA at delivery and has increasingly been used as the reference standard in studies evaluating the accuracy of dating ultrasound and other methods of estimating GA.

We have recently developed statistical models that can estimate GA postnatally using metabolic markers from heel-prick blood samples collected shortly after birth in combination with clinical and demographic data, and externally validated these models with data from low- and middle-income countries. Although we have been able to demonstrate high accuracy (±7 days) relative to first-trimester ultrasound as the reference standard, we hypothesized that in newborns conceived using ART, a GA reference standard derived from the age of embryo and DET would have the highest possible accuracy. Our objectives in this study were to validate both first-trimester ultrasound GA dating and postnatal metabolomic GA estimation models in a cohort of newborns conceived via ART, both overall and in important subgroups of interest (preterm birth, small for GA [SGA], and multiple birth).

Materials and Methods

Study design and data sources

This was a retrospective database cohort study approved by the Ottawa Health Science Network Research Ethics Board (20140724-01H) and the Children’s Hospital of Eastern Ontario Research Ethics Board (15-143X) on November 5, 2015. Ontario, the most populous Canadian province, has approximately 140,000 live births annually. Virtually all newborns (>99%) undergo expanded newborn screening, and the data are captured in the Better Outcomes Registry & Network (BORN) provincial birth registry. From the BORN registry, we identified a cohort of infants born between January 2015 and December 2017. For infants to be included in the cohort, the following data needed to be available: (1) GA measured by first-trimester dating ultrasound, (2) clinical covariates, and (3) a complete newborn screening panel of laboratory results for blood samples collected within 48 hours of birth. Within this cohort, assisted conceptions resulting from fresh embryo transfers were identified using the Canadian Assisted Reproductive Technologies Registry Plus (CARTR+) database, a registry of patient data from clinics providing assisted reproduction services in Canada, which is maintained at BORN. The Ontario CARTR+ data have been individually linked to the Ontario maternal newborn registry data. ART conceptions were set aside from the overall cohort, and then the data for spontaneous conceptions were randomly partitioned into a model derivation sample (80%) that was used for model development and a test sample (20%) that was used to internally validate model performance in births from spontaneously conceived pregnancies. The ART conceptions were then used as an additional independent test cohort. Treatment cycles from stored frozen embryos were excluded because of data quality issues in the reporting of date and stage information for each thawed embryo. Figure 1 describes the exclusions made during cohort creation.

For the ART test cohort, in addition to the reference GA available via first-trimester dating ultrasound, we calculated an additional reference GA based on DET and the stage of development of the embryo (in days) at transfer using the formula:

\[ \text{GA (days)} = \text{(date of birth − embryo transfer date)} + \text{embryo stage (days)} + 14. \]

Sample collection and analysis of metabolic markers used in gestational age algorithm development. Heel-prick blood samples are collected within the first 48 hours after birth from virtually all newborns in the province (>99%) as part of Ontario’s universal expanded newborn screening program. Samples were all analyzed centrally at the Newborn Screening Center at the Newborn Screening Ontario newborn screening program.
Ontario laboratory in Ottawa, Ontario, Canada. The analytes available as candidate predictors in GA estimation models are all measured as part of the newborn screening program, and include hemoglobin profiles, amino acids, acylcarnitines, endocrine and immunity markers, enzymes, and coenzymes (Table 1).

**Model development.** All analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC) and R 3.3.2 (R Core Team, Vienna, Austria). Models were developed using clinical covariates and laboratory biomarker results from heel-prick blood samples in the Ontario spontaneous conception cohort using ultrasound-based GA as the reference GA. Data preparation steps, including standardization and log transformation were applied independently in both the Ontario spontaneous conception cohort and in the assisted reproduction cohort. Models were fit and all parameters estimated using the model training cohort, and model performance was evaluated in the ART and spontaneous conception test cohorts, which were not used.
TABLE 1
Newborn screening analytes included in predictive models

| Hemoglobins               | Adult hemoglobin: HbA(A) | Fetal hemoglobin: HbF (F), acetylated HbF (F1) |
|---------------------------|--------------------------|-----------------------------------------------|
| Endocrine markers         | 17-hydroxyprogesterone (17-OHP), thyroid-stimulating hormone (TSH) |
| Amino acids               | Arginine (arg); phenylalanine (phe); alanine (ala); leucine (leu); ornithine (orn); citrulline (cit); tyrosine (tyr); glycine (gly); methionine (met); valine (val) |
| Acylcarnitines            | C0; C2; C3; C4; C5; C5:1; C6; C8; C8:1; C10; C10:1; C12; C12:1; C14; C14:1; C14:2; C16; C18; C18:1; C18:2; C10:1; C12:1; C14:1; C14:2; C4OH; C5:1; C5DC; C5OH; C6DC; C16:0H; C16:1OH; C18OH; C18:1OH; C3DC; C4DC |
| Enzyme markers            | Biotinidase; immunoreactive trypsinogen (IRT) |
| Immune markers            | T-cell receptor excision circles (TREC) |

In previously published work, we also fit models including only sex, multiple gestation, and birthweight, which were less accurate; hence, we only considered the above models in the current study. Through partial Spearman correlation analysis, the strength of association of each analyte, birthweight, sex, and multiple gestation with GA was determined. Seven analytes and birthweight were most strongly associated with GA after adjusting for correlation with all other analytes and clinical covariates in the partial Spearman correlation analysis. These 7 analytes (17-OH progesterone, complement component 5, tyrosine, C4DC, C5DC, T-cell receptor excision circles, adult/fetal hemoglobin ratio) and birthweight were assigned additional degrees of freedom by including them in the model as spline effects using restricted cubic splines with 5 knots to allow for nonlinearity.

Multivariable linear regression using elastic-net regularization was used to model the large number of covariates, cubic spline effects, and pairwise interactions included in each model. Elastic net combines 2 types of penalization, blending ridge regression and LASSO-type penalties. Final model equations were then used to estimate GA in the test subset of the Ontario spontaneous conception cohort and in the ART cohort. Details of data preparation and model training are provided in the Appendix.

Model validation. The accuracy of GA estimates from dating ultrasound and from metabolomic models was assessed by comparing the estimated GA with reference GA in the independent test cohorts. Model performance of our metabolomic GA models has been reported previously in Ontario infants including both spontaneous and ART conceptions. Results in the test cohort of exclusively spontaneous conceptions were very similar to those reported previously, but were included in the study results to facilitate comparison of model performance between ART and spontaneous conceptions. Metabolomic model estimates and dating ultrasound estimates were compared with DET reference GA. Metabolomic model estimates were also validated against ultrasound reference GA. To quantify the accuracy of each estimated GA, we calculated 2 measures: the mean absolute error (MAE) and the square root of mean squared error (RMSE), which are both in the same units of the outcome (GA in weeks). Lower values of MAE and RMSE reflect higher accuracy, and both are reported to facilitate comparison with other reported algorithm validations. We also calculated the percentage of infants with GA correctly estimated within 7 days of reference GA. We assessed model performance overall and in important subgroups, namely preterm birth (<37 weeks’ gestation), multiple birth, and SGA <10th percentile (SGA10) within categories of gestational week at delivery and infant sex based on International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21) standards. We calculated 95% bootstrap percentile confidence intervals (CIs) based on 500 bootstrap replicates for all validation performance metrics. In addition, we estimated the preterm birth rate by calculating the proportion of model-based GA estimates that were <37 weeks, and compared these with the observed preterm birth rate in each cohort (based on either ultrasound or DET reference GA).

Results
Participant characteristics
The Ontario reference cohort was partitioned into training and test samples. The model derivation cohort included N=202,300 spontaneous conceptions, and the testing cohorts included N=50,735 spontaneous conceptions and N=1924 ART conceptions from fresh embryo transfers. Characteristics of infants in the Ontario reference and assisted reproduction cohorts are presented in Table 2. In the spontaneous conceptions test cohort, prevalence of preterm gestation (GA <37 weeks) based on ultrasound GA was 7.0%, and 4.0% of infants were classified as SGA. In the ART cohort, the rate of preterm gestation was 13.5% according to ultrasound GA,
and 5.8% of infants were classified as SGA. Compared with the Ontario spontaneous cohort, newborns in the ART cohort were more likely to be female (55.0% vs 48.7%), of low birthweight (<2500 g; 13.3% vs 5.7%), or from a multiple birth (8.7% vs 1.0%).

### Performance of ultrasound

Overall, first-trimester dating ultrasound was accurate to within approximately ±1.5 days compared with DET reference GA in the ART cohort (MAE, 0.21 [95% CI, 0.20–0.23]). Accuracy was best in term births and moderately lower in SGA, preterm, and multiple births, but still within ±2 days of DET reference GA on average (Table 3).

**Performance of the metabolic gestational age model in Ontario spontaneous conception and assisted reproduction cohorts.** Details regarding the fitted models are presented in the Appendix. In general, Model 2, which included sex, multiple gestation, birthweight, and metabolomic predictors, provided more accurate GA estimates compared with Model 1 (Table 4). Model 2 was accurate to within 5 days of ultrasound-assigned values when applied to the spontaneous conception cohort (MAE, 0.70 [95% CI, 0.69–0.70] weeks), with similar accuracy in the ART cohort (MAE, 0.70 [95% CI, 0.68–0.73] weeks) compared with ultrasound GA. When GA based on DET was used as the reference GA, the accuracy of model estimates slightly improved (MAE, 0.68 [95% CI, 0.66–0.71] weeks). Accuracy of GA estimates in all models was highest in term infants and tended to be lower in preterm and SGA infants, but Model 2 consistently provided the most accurate GA estimates in both term and preterm infants. GA was correctly estimated to within 1 week of ultrasound-assigned values for 75.7% (95% CI, 75.4–76.1) of spontaneous conceptions, and for 74.6% (95% CI, 72.7–76.5) and 75.4% (95% CI, 73.7–77.3) of conceptions in the ART cohort compared with ultrasound and DET reference GA, respectively. Model 2 estimated GA within 2 weeks in >95% of infants in both the spontaneous and ART cohorts overall.

Model 2 outperformed Model 1 when applied to samples from preterm infants from spontaneous and assisted conceptions. When applied to samples from preterm infants from spontaneous conceptions, Model 2 correctly estimated GA to within 6 days of ultrasound-assigned values (MAE, 0.91 [95% CI, 0.88–0.93] weeks), and performed markedly better than Model 1 (MAE, 1.13 [95% CI, 1.10–1.15] weeks). In the ART cohort, Model 2 was more accurate than Model 1, and estimated GA to within approximately 6 days on average in preterm infants (MAE, 0.82 [95% CI, 0.73–0.91] weeks) compared with ultrasound GA.

### Table 2

**Infant characteristics**

| Characteristics          | Infants conceived spontaneously (n=50,735) | Infants conceived with ART (n=1924) |
|--------------------------|-------------------------------------------|------------------------------------|
| Sex, n (%)               |                                           |                                    |
| Male                     | 26,021 (51.3%)                            | 866 (45.0%)                        |
| Female                   | 24,714 (48.7%)                            | 1058 (55.0%)                       |
| Gestational age category (wk), n (%) |                                      |                                    |
| ≥37                      | 47,192 (93.0%)                            | 1664 (86.5%)                       |
| 32–36                   | 3209 (6.3%)                               | 230 (12.0%)                        |
| 28–31                   | 247 (0.5%)                                | 22 (1.1%)                          |
| <28                      | 87 (0.2%)                                 | 8 (0.4%)                           |
| Low birthweight (<2500 g), n (%) |                                       |                                    |
| Small for gestational age (SGA10), n (%) |                                    |                                    |
| Multiple birth, n (%)    | 1498 (3.0%)                               | 260 (13.5%)                        |

ART, assisted reproductive technology; SGA10, small for gestational age <10th percentile.

Hawken. Gestational age determination in assisted conceptions. Am J Obstet Gynecol Glob Rep 2022.

### Table 3

**Validation of first-trimester dating ultrasound in assisted reproductive technology conceptions compared with a reference standard based on date and stage of embryo at transfer**

| Assisted conceptions, ultrasound GA (date of embryo transfer reference GA) | Overall | <37 wk | SGA10 | Multiple birth |
|--------------------------------------------------------------------------|---------|--------|-------|---------------|
| N                                                                        | N=1924  | N=260  | N=112 | N=260         |
| MAE (95% CI)                                                             | 0.21    | 0.26   | 0.27  | 0.24          |
| (0.20–0.23)                                                              | (0.22–0.31) | (0.17–0.40) | (0.20–0.28) |
| RMSE (95% CI)                                                            | 0.37    | 0.47   | 0.64  | 0.37          |
| (0.32–0.43)                                                              | (0.32–0.68) | (0.29–1.01) | (0.30–0.46) |
| % ±1 wk (95% CI)                                                         | 99.2    | 98.5   | 96.4  | 98.5          |
| (98.8–99.6)                                                              | (96.9–99.6) | (92.4–99.2) | (96.9–99.6) |

CI, confidence interval; GA, gestational age; MAE, mean absolute error; RMSE, square root of mean squared error; SGA10, small for gestational age <10th percentile.

Hawken. Gestational age determination in assisted conceptions. Am J Obstet Gynecol Glob Rep 2022.
Table 4: Validation of metabolomic model estimated gestational age compared with: (1) first-trimester dating ultrasound reference gestational age in the spontaneous conception cohort, (2) first-trimester dating ultrasound in the assisted reproductive technology cohort, and (3) date and stage of embryo at transfer reference standards in the assisted reproductive technology cohort.

| Model 1: sex, multiple gestation, analytes (no birthweight) | Model 2: sex, multiple gestation, analytes+birthweight |
|-----------------------------------------------------------|-------------------------------------------------------|
| 1. Spontaneous conceptions (ultrasound reference GA)      |                                                       |
| N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) | N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) |
| N=50,735     | 0.79 (0.79–0.80) | 1.13 (1.10–1.15) | 69.8 (69.3–70.2) | N=50,735     | 0.70 (0.69–0.70) | 1.02 (1.01–1.02) | 69.8 (69.3–70.2) |
| N=3543       | 1.02 (1.01–1.02) | 1.42 (1.38–1.45) | 52.1 (50.5–53.6) | N=3543       | 1.13 (1.10–1.15) | 1.42 (1.38–1.45) | 52.1 (50.5–53.6) |
| N=2036       | 1.15 (1.11–1.19) | 1.44 (1.38–1.50) | 62.9 (60.7–64.9) | N=2036       | 1.15 (1.11–1.19) | 1.44 (1.38–1.50) | 62.9 (60.7–64.9) |
| N=1498       | 1.15 (1.11–1.19) | 1.44 (1.38–1.50) | 65.4 (63.0–67.6) | N=1498       | 1.15 (1.11–1.19) | 1.44 (1.38–1.50) | 65.4 (63.0–67.6) |

2. ART conceptions (ultrasound reference GA)

| N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) | N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) |
| N=1924        | 0.81 (0.78–0.84) | 1.11 (1.01–1.22) | 69.0 (66.8–71.1) | N=1924        | 0.79 (0.76–0.81) | 1.06 (0.96–1.15) | 69.0 (66.8–71.1) |
| N=260         | 1.05 (1.00–1.09) | 1.44 (1.29–1.58) | 55.4 (49.3–61.2) | N=260         | 1.02 (0.97–1.03) | 1.34 (1.22–1.46) | 55.4 (49.3–61.2) |
| N=112         | 0.96 (0.80–1.13) | 1.29 (1.07–1.54) | 65.1 (55.7–74.2) | N=112         | 1.00 (0.97–1.03) | 1.34 (1.22–1.46) | 65.1 (55.7–74.2) |
| N=260         | 0.84 (0.75–0.92) | 1.10 (0.98–1.21) | 69.3 (63.4–75.1) | N=260         | 0.81 (0.74–0.86) | 1.06 (0.96–1.17) | 69.3 (63.4–75.1) |

3. ART conceptions (date of embryo transfer reference GA)

| N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) | N             | MAE (95% CI)   | RMSE (95% CI)  | % ± 1 wk (95% CI) |
| N=1924        | 0.79 (0.76–0.81) | 1.06 (0.96–1.15) | 70.2 (68.1–72.2) | N=1924        | 0.79 (0.76–0.81) | 1.06 (0.96–1.15) | 70.2 (68.1–72.2) |
| N=260         | 1.00 (0.97–1.03) | 1.34 (1.22–1.46) | 55.1 (49.2–60.9) | N=260         | 1.00 (0.97–1.03) | 1.34 (1.22–1.46) | 55.1 (49.2–60.9) |
| N=112         | 0.89 (0.76–1.03) | 1.14 (0.98–1.30) | 66.7 (57.6–75.3) | N=112         | 0.89 (0.76–1.03) | 1.14 (0.98–1.30) | 66.7 (57.6–75.3) |
| N=260         | 0.81 (0.74–0.90) | 1.06 (0.96–1.17) | 70.8 (65.2–76.2) | N=260         | 0.81 (0.74–0.90) | 1.06 (0.96–1.17) | 70.8 (65.2–76.2) |

ART, assisted reproductive technology; CI, confidence interval; GA, gestational age; MAE, mean absolute error; RMSE, square root of mean squared error; SGA10, small for gestational age <10th percentile.

Hawken. Gestational age determination in assisted conceptions. Am J Obstet Gynecol Glob Rep 2022.
sound reference GA, and accuracy was higher than that of DET reference GA (MAE, 0.78 [95% CI, 0.71–0.86] weeks).

When applied to SGA10 infants from both spontaneous and ART conceptions, Model 1, which excluded birthweight, was slightly more accurate than Model 2. Model 1’s accuracy was highest in the ART cohort when compared with DET reference GA (MAE, 0.89 [95% CI, 0.76–1.03] weeks vs 0.96 [95% CI, 0.80–1.13] weeks and 0.93 [95% CI, 0.89–0.96] weeks in ART and spontaneous conceptions, respectively, using ultrasound reference GA).

The accuracy of Models 1 and 2 in infants from multiple births was slightly diminished compared with overall model performance in infants from spontaneous conceptions, and in infants from ART conceptions with either ultrasound or DET reference GA. However, it was higher than the accuracy reached for both SGA10 and preterm infants in all cases. Model 2 was the most accurate in infants from multiple births (MAE, 0.78 [95% CI, 0.75–0.81] in spontaneous conceptions [Table 3]; MAE, 0.68 [95% CI, 0.61–0.74] in ART conceptions with ultrasound reference GA; and MAE, 0.65 [95% CI, 0.59–0.72] with DET reference GA [Table 4]).

Scatter plots of DET reference GA vs estimated GA from ultrasound and estimated GA from metabolomic Models 1 and 2 in the ART cohort are presented in Figure 2.

**Estimated preterm birth rate**
In the spontaneous conception cohort, the preterm birth prevalence was 7.0% according to ultrasound reference GA. Model 2, which demonstrated the highest accuracy overall, estimated the preterm birth proportion to be 7.3% (95% CI, 7.1–7.6) in the spontaneous conception test cohort. In the ART cohort, the preterm birth prevalence was 13.5% according to ultrasound reference GA, and 13.6% according to DET. Model 2 estimated the prevalence to be 14.9% (95% CI, 13.3–16.5).

**Comment**
**Principal findings**
We investigated the accuracy of first-trimester dating ultrasound and postnatal metabolomic GA estimation algorithm in a population of infants conceived with ART. We assumed that DET would provide the most accurate reference GA possible because previous studies have reported that first-trimester dating ultrasound may have a margin of error of ±3 to 8 days or more, depending on the reference charts used in conjunction with crown–rump length measurements. When first-trimester ultrasound was compared with GA derived from date of transfer and stage of the embryo in infants conceived with ART, we found that ultrasound was accurate to within approximately 1.5 days of DET reference GA (MAE, 0.21 [95% CI, 0.20–0.23] weeks). Accuracy of first-trimester ultrasound was decreased in preterm and SGA infants, where MAE was 0.26 (0.22–0.31) and 0.27 (0.17–0.40) weeks, respectively, still within ±2 days of DET reference GA. When compared with reference GA derived from DET, we found that the metabolomic GA estimation models were still accurate to within approximately ±5 days of DET reference GA overall (MAE, 0.79 [0.76–0.81] weeks). When ultrasound was used as the reference GA in validating the metabolomic model, the MAE was slightly higher overall (MAE, 0.81 [0.78–0.84] weeks). In general, the accuracy of GA estimates derived from ultrasound or metabolomic models was highest in term infants and lower in preterm and SGA newborns. We observed a small but consistent trend across validation metrics of higher apparent metabolomic model accuracy when DET reference GA was used compared with when ultrasound reference GA was used, and this improvement in accuracy was larger in magnitude in preterm and SGA infants. Although less accurate than reference dating ultrasound, our metabolomic models provided very good estimates of the preterm birth rate at the population level in both the spontaneous cohort (7.3% for Model 2 vs 7.0% via ultrasound) and in Figure 2.
the ART cohort (14.9% for Model 2 vs 13.5% via ultrasound and 13.6% via DET reference GA).

**Clinical implications**

In interpreting our results, it is important to note that the dating time point for ultrasound is much closer to conception than the postnatal biological sample used in the metabolomic dating algorithms, which may have contributed to the observed accuracy. Furthermore, the metabolic markers are derived from the developmental stage of the delivered infant, which may be more or less developmentally mature than any specific GA might suggest. This can be seen in the wide variability in clinical presentation in preterm infants of the same GA. It is also important to consider that metabolic markers measured in preterm infants could be affected by any pathophysiology underlying the preterm birth.

**Research implications**

Our study reinforces the accuracy of first-trimester ultrasound for gestational dating and demonstrates the promise of metabolomics biomarkers in the absence of other methods of accurate gestational dating. We evaluated the use of the GA model in low-resource settings to estimate population levels of preterm birth and found the model to be accurate to within ±6 days. A health economic study estimated that the annual incremental cost to implement this metabolic algorithm would be approximately $100,000. Although this would clearly be cost-prohibitive for many low-resource settings, it could become more accessible if international funding bodies and/or governments prioritize the implementation of such technologies to improve estimates of the burden of preterm birth. Evolving technology and economies of scale could bring incremental costs down substantially.

**Strengths and limitations**

Although we demonstrated the potential advantages of validating our models against DET reference GA, there were key limitations that may have affected the generalizability of our findings. First, women who undergo ART are different, at the population level, in age, fertility history, and socioeconomic status. Second, people experiencing infertility who seek ART treatment are a self-selected population. There are also concerns inherent to the ART process itself, such as the unclear effects of ART laboratory techniques on fetal growth (eg, effects of laboratory media in which embryos are grown) and the potential lag time between ovulation and conception. Our ART cohort was also limited to live births resulting from fresh embryo transfers because data quality was unacceptably low in embryo transfers from thawed, frozen embryos. We were also limited to births for which we could confirm that an ultrasound-based GA estimate was available in addition to GA estimated from DET, and there was no information available about whether the ultrasound GA determination was blinded from the DET GA determination. We also observed a much higher preterm gestation, SGA, and multiple gestation prevalence in our ART test cohort than in the spontaneous conception cohort. Therefore, model performance metrics calculated in the ART test cohort may not be representative of expected model performance in a population cohort. We have, however, demonstrated the utility of our approach in comparing relative performance of different models, and different reference standards. This has potential to allow comparative validation of competing candidate models against the most accurate reference standard GA available.

**Conclusions**

We demonstrated that ultrasound dating matches closely to the exact date of conception based on ART. We found that the metabolomic GA estimation models may be slightly more accurate according to DET reference GA than previously reported model validations based on ultrasound reference GA. This improvement was relatively small in magnitude (ie, the improvement was a small fraction of a day in most cases) and may not represent a clinically meaningful difference, although it was consistently observed across almost all metrics and subpopulations of interest. Overall, estimated preterm birth rates were similar across all 3 estimation methods, with the metabolomic model slightly overestimating the prevalence of preterm gestation compared with ultrasound estimates. Preterm birth rates in the ART cohort were nearly double the rates in the spontaneous cohort.

Our findings support the accuracy of ultrasound as a GA dating tool. They also support the use of metabolic GA dating algorithm in settings where ultrasound is not available, such as in low- and middle-income countries and in other low-resource settings; however, the margin of error is increased compared with ultrasound.

**Supplementary materials**

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.xagr.2022.100091.
9. Knight B, Bereton A, Powell RJ, Liversedge H. Assessing the accuracy of ultrasound estimation of gestational age during routine antenatal care in in vitro fertilization (IVF) pregnancies. Ultrasound 2018;26:49–53.

10. Dias T, Mahsud-Doran S, Thilaganathan B, Papageorghiou A, Bhide A. First-trimester ultrasound dating of twin pregnancy: are singleton charts reliable? BJOG 2010;117:979–84.

11. Gjerris AC, Loft A, Pinborg A, Tabor A, Christiansen M. First-trimester screening in pregnancies conceived by assisted reproductive technology: significance of gestational dating by oocyte retrieval or sonographic measurement of crown-rump length. Ultrasound Obstet Gynecol 2008;32:612–7.

12. Wilson K, Hawken S, Potter BK, et al. Accurate prediction of gestational age using newborn screening analyte data. Am J Obstet Gynecol 2016;21:513.e1–9.

13. Wilson K, Hawken S, Murphy MSQ, et al. Postnatal prediction of gestational age using newborn fetal hemoglobin levels. EBioMedicine 2017;15:203–9.

14. Murphy MS, Hawken S, Cheng W, et al. External validation of postnatal gestational age estimation using newborn metabolic profiles in MATLAB. Bangladesh. Elife 2019;8:e42627.

15. Hawken S, Murphy MSQ, Ducharme R, et al. External validation of machine learning models including newborn metabolomic markers for postnatal gestational age estimation in East and South-East Asian infants. Gates Open Res 2021;4:164.

16. Túñón K, Elk-Nes SH, Grettum P, Von Düring V, Kahn JA. Gestational age in pregnancies conceived after in vitro fertilization: a comparison between age assessed from oocyte retrieval, crown-rump length and biparietal diameter. Ultrasound Obstet Gynecol 2000;15:41–6.

17. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. New York, NY: Springer; 2001.

18. Villar J, Cheikh Ismail LC, Victora CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. Lancet 2014;384:257–68.

19. Mahendru AA, Wilhelm-Benartzi CS, Wilkinson IB, McEniery CM, Johnson S, Lees C. Gestational length assignment based on last menstrual period, first trimester crown-rump length, ovulation, and implantation timing. Arch Gynecol Obstet 2016;294:867–76.

20. Chalouhi GE, Bernard JP, Benoist G, Nasr B, Ville Y, Salomon LJ. A comparison of first trimester measurements for prediction of delivery date. J Matern Fetal Neonatal Med 2011;24:51–76.

21. McLennan AC, Schluter PJ. Construction of modern Australian first trimester ultrasound dating and growth charts. J Med Imaging Radiat Oncol 2008;52:471–9.

22. McLennan AC, Schluter PJ. Construction of modern Australian first trimester ultrasound dating and growth charts. J Med Imaging Radiat Oncol 2008;52:471–9.

23. Coyle K, Quan AML, Wilson LA, et al. Cost-effectiveness of a gestational age metabolic algorithm for preterm and small-for-gestational-age classification. Am J Obstet Gynecol MFM 2021;3:100279.

24. Napolitano R, Dhami J, Ohuma EO, et al. Pregnancy dating by fetal crown-rump length: a systematic review of charts. BJOG 2014;121:556–65.

25. Dumoulin JC, Land JA, Van Montfoort AP, et al. Effect of in vitro culture of human embryos on birthweight of newborns. Hum Reprod 2010;25:605–12.

26. Nelissen EC, Van Montfoort AP, Coonen E, et al. Further evidence that culture media affect perinatal outcome: findings after transfer of fresh and cryopreserved embryos. Hum Reprod 2012;27:1966–76.

27. Santos MA, Kuik EW, Macklon NS. The impact of ovarian stimulation for IVF on the developing embryo. Reproduction 2010;139:23–34.

28. McLennan AC, Schluter PJ. Construction of modern Australian first trimester ultrasound dating and growth charts. J Med Imaging Radiat Oncol 2008;52:471–9.

29. Choux C, Carmignac V, Bruno C, Sagot P, Varian D, Fauque P. The placenta: phenotypic and epigenetic modifications induced by assisted reproductive technologies throughout pregnancy. Clin Epigenetics 2015;7:87.