Received: 7 August 2023 | Revised: 24 November 2023 | Accepted: 28 November 2023

DOI: 10.1111/ppe.13029

ORIGINAL ARTICLE

Paediatric and Perinatal Epidemiology WILEY

# Unveiling sex bias and adverse neonatal outcomes in ultrasound estimation of gestational age: A population-based cohort study

Anders Einum[1,2] | Linn Marie Sørbye[2,3] | Roy Miodini Nilsen[3] | Cathrine Ebbing[1,2] | Nils-Halvdan Morken[1,2]

[1]Department of Clinical Science, University of Bergen, Bergen, Norway

[2]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway

[3]Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway

**Correspondence**
Anders Einum, Department of Clinical Science, University of Bergen, Bergen, Norway.
Email: anders.einum@uib.no

**Funding information**
Gidske and Peter Jacob Sørensens fund; Helse Vest; Odd Fellow Medical Fund

## Abstract

**Background:** Gestational age estimation by second-trimester ultrasound biometry introduces systematic errors due to sex differences in early foetal growth, consequently increasing the risk of adverse neonatal outcomes. Ultrasound estimation earlier in pregnancy may reduce this bias.

**Objectives:** To investigate the distribution of sex ratio by gestational age and estimate the risk of adverse outcomes in male foetuses born early-term and female foetuses born post-term by first- and second-trimester ultrasound estimations.

**Methods:** This population-based study compared two cohorts of births with gestational age based on first- and second-trimester ultrasound in the Medical Birth Registry of Norway between 2016 and 2020. We used a log-binomial regression model to estimate adjusted relative risk (RR) with 95% confidence interval (CI) for Apgar score <7 at 5 min, umbilical artery pH <7.05, neonatal intensive care unit (NICU) admission and respiratory morbidity in relation to foetal sex.

**Results:** The sex ratio at birth in gestational weeks 36–43 showed less male predominance in pregnancies estimated in first compared to second trimester. Any adverse outcome was registered in 627 of 4470 male infants born in gestational weeks 37–38 and 618 of 6406 females born ≥41 weeks. Male infants born in weeks 37–38 had lower risk of NICU admission (RR 0.76, 95% CI 0.58, 0.99), Apgar score <7 at 5 min (RR 0.63, 95% CI 0.28, 1.41) and respiratory morbidity (RR 0.68, 95% CI 0.37, 1.25) in first- compared to second-trimester estimations. Female infants estimated in first trimester born ≥41 weeks had lower risk of umbilical artery pH <7.05, NICU admissions and respiratory morbidity; however, CIs were wide.

**Conclusions:** Early ultrasound estimation of gestational age may reduce the excess risk of adverse neonatal outcomes and highlight the role of foetal sex and the timing of ultrasound assessment in the clinical evaluation of preterm and post-term pregnancies.

KEYWORDS
crown–rump length, foetal development, gestational age, measurement error, sex ratio, ultrasound

## 1 | BACKGROUND

The exact time of human fertilization as the basis of calculating gestational age is rarely known. Except for pregnancies conceived by assisted reproductive technology (ART), where gestational age is based on the date of embryo transfer to the uterus, the time of fertilization is stipulated in retrospect.

In modern pregnancy care, ultrasound biometry preponderates over the last menstrual period (LMP) to estimate gestational age. The biometry is based on standardized sonographic measurements of foetal anatomy, most frequently biparietal diameter or head circumference in the second trimester or CRL (crown–rump length) in the first trimester.[1] Ultrasound is widely recommended as it is more reliable in estimating the date of birth and reduces the proportion of post-term births compared to LMP in population studies.[2,3] However, it is important to be aware of the limitations of the method as foetal size does not always equal age. The method assumes homogenous foetal growth velocities preceding the examination and is a simplification of biological variation. In addition to early growth restriction, physiologic, genetic and environmental factors may influence early embryonic development and growth.[4–7] Thus, systematic misclassification of gestational age by ultrasound is possible.

The clinical management of all pregnancies relies on the accurate estimation of gestational age. Preterm and post-term pregnancies are associated with a higher risk of foetal morbidity and mortality and are subject to different foetal monitoring, induction of labour and obstetrical intervention than term births, and therefore correct classification is vital. Because sex influences foetal growth velocity in early pregnancy, larger male foetuses are estimated to be older than female foetuses by second-trimester ultrasound biometry.[8,9] By systematically underestimating remaining pregnancy length in males, ultrasound dating results in male predominance in post-term births, as opposed to more equal sex distribution by gestational age in estimations by LMP which is not affected by differences in growth.[10] This bias implies that female foetuses in post-term pregnancies may be more severely post-term than their male counterparts, and male foetuses born early-term may be more preterm than females at the same gestational age. Research has shown an excess risk of adverse neonatal outcomes in early-term males and post-term females with second-trimester ultrasound age assessment compared with LMP, supporting this hypothesis.[10,11]

International guidelines recommend ultrasound CRL measurements in the first trimester for estimation of gestational age as it has lower random and systematic errors than other biometric parameters.[1,12] Also, the impact of sex-dependent divergence in growth trajectories is reduced when measurements are performed earlier in pregnancy.[13] How these features of early ultrasound may impact the adverse outcomes in early-term males and post-term females reported in second-trimester ultrasound has not been studied to our knowledge, and may provide valuable information in the clinical assessment of risk pregnancies. We aimed to compare sex ratios and investigate differences in adverse neonatal outcomes in early-term male and post-term female births in first- versus second-trimester ultrasound estimation of gestational age.

### SYNOPSIS

#### Study question

Investigate the distribution of sex ratio by gestational age and whether sex-related adverse neonatal outcomes differ by first- and second-trimester ultrasound estimations.

#### What's already known

Compared to estimation by the last menstrual period, gestational age estimation by second-trimester ultrasound introduces systematic errors due to sex differences in early foetal growth. Consequently, the sex ratio at birth is biased, and the risk of adverse neonatal outcomes increases in male foetuses born early term and female foetuses born post-term.

#### What this study adds

Gestational age assessment by ultrasound obtained in the first trimester may reduce the excess risk of adverse neonatal outcomes compared to second-trimester estimations and highlights the clinical importance of foetal sex and estimation method in preterm and post-term gestations.

## 2 | METHODS

### 2.1 | Study population

The Medical Birth Registry of Norway (MBRN) contains information on all births in the country by mandatory collection and is the main data source of the study.[14] The mothers' unique national identity numbers were used to link data on attained education and maternal country of birth from Statistics Norway. The study population was restricted to singleton births between gestational weeks 22+0 and 45+0 in the period 2016–2020, where data on different ultrasound methods were recorded in the MBRN. We excluded pregnancies with foetal malformations, induced abortions and extreme misclassifications defined as birthweight ≥4 standard deviations from mean birthweight by gestational age. Two independent cohorts of ultrasound estimation were defined for the calculation of sex ratios by gestational age. The sex ratio in ART pregnancies was calculated as a third cohort for comparison, but not included in the analysis of other outcomes (Figure 1).

Measurements of head circumference and biparietal diameter are considered equal methods for estimating gestational age in the second trimester,[15] and both methods were included in the

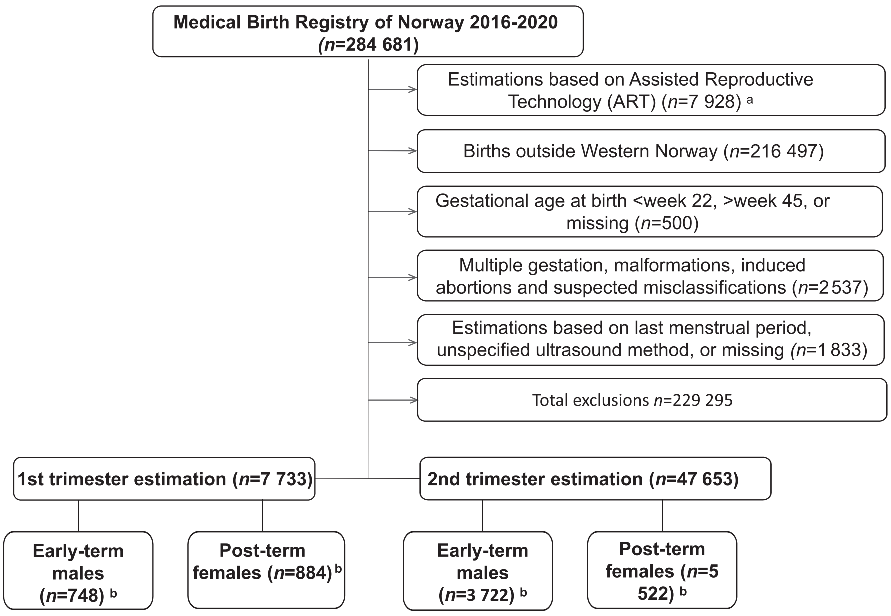


**FIGURE 1** Flowchart of study population. [a]Population included in the calculation of sex ratios. [b]Cohorts defined for the analysis of adverse neonatal outcomes.

second-trimester cohort. First-trimester CRL measurements were obtained either in gestational week 12 by public healthcare examiners certified by the Fetal Medicine Foundation (FMF), or in weeks 8–12 by private healthcare providers. In the latter case, the measurement quality was scrutinized by an FMF-certified examiner before they were accepted as gestational age estimates. We identified MBRN registrations from Western Norway using CRL measurements to comprise the first-trimester estimation cohort and consequently restricted both ultrasound cohorts to Western Norway to avoid geographical variation.

In the analysis of adverse neonatal outcomes, we restricted the population to early-term male births at gestational age from 37+0 to 38+6 weeks following previous research, and post-term female births at ≥41+0 weeks in line with the national post-term induction policy in the study period.[11,16] In two secondary analyses, we investigated the same outcomes in cohorts of early-term females and post-term males as a proof of concept and stratified on iatrogenic and spontaneous birth to explore whether clinical intervention affected the findings.

## 2.2 | Outcome measures

Sex ratio was calculated as the number of male infants born/number of female infants born per gestational week in cohorts of ART, first-trimester ultrasound and second-trimester ultrasound estimation respectively. We analysed four adverse outcomes coded as dichotomous variables indicative of infant compromise. *Apgar score* is the clinician's evaluation of neonatal vitality at 1, 5 and 10 min after birth on a scale from 0 to 10, where a score <7 at 5 min is associated with an increased risk of neonatal mortality and morbidity.[17–19] *Umbilical artery pH* obtained from blood sampling immediately after delivery was dichotomized with values <7.05 defined as birth asphyxia. Subjects with only an arterial sample or a difference between arterial and venous pH <0.02 were not included in the analysis due to misclassification or samples drawn from the same cord vessel respectively. Norwegian guidelines recommend obtaining cord blood in all births,[20] and the substantial number of missing values are believed to be related to technical issues in sampling, thus missing at random. As optimal clinically relevant cutoff values are debated,[21,22] we also reported median values with interquartile range for Apgar score and mean values with standard deviations for umbilical artery pH for comparison. *Admission to a neonatal intensive care unit* (NICU) indicates a need for paediatric observation and treatment. *Respiratory morbidity* was defined as neonatal respiratory distress syndrome (RDS) based on typical symptoms and findings on chest X-ray, or assisted ventilation by either continuous positive airway pressure (CPAP) or respirator. A *composite outcome* was defined as having any of the outcomes Apgar score <7 at 5 min, umbilical artery pH <7.05, NICU admission and/or respiratory morbidity.

## 2.3 | Covariates

The Norwegian public healthcare system provides free antenatal care including ultrasound estimation of gestational age in the second trimester between gestational weeks 18–20. However, additional ultrasound examinations in the first trimester by private healthcare providers are common,[23] and their use might be subject to selection differences by socioeconomic position associated with neonatal health outcomes.[24] Maternal smoking is well documented to restrict foetal growth, bias ultrasound estimation and thereby gestational age at birth.[4] The access to first-trimester prenatal diagnostics in Norway during the study period was very restricted in an international context, provided only on the indications of advanced maternal age, chronic maternal illness, previous foetal malformations and chromosome disorders, complicated obstetric history or severe anxiety. As these indications are also independent risk factors for adverse neonatal outcomes, any first-trimester estimation obtained as part of prenatal diagnostics might confound the association by indication. Thus, maternal age, chronic maternal illness, smoking, maternal country of birth and level of education (a proxy for socioeconomic position) were accounted for in the adjusted analysis.

### 2.4 | Statistical analysis

Covariates possibly confounding the association in early-term males and post-term females were identified by drawing a directed acyclic graph and included in the model (Figure S1).[25,26] To investigate differences in birth outcomes between the cohorts, we used log-binomial regression models (with log-link function) to estimate unadjusted and adjusted relative risks (RR) with a 95% confidence interval (CI). To account for intra-individual correlation for mothers contributing with more than one birth in the cohorts, we used a robust variance estimation of model parameters. In all analyses, the second-trimester cohort was used as a reference.

Statistical analyses were performed using the Stata release 17.0 software package (StataCorp LLC).

### 2.5 | Missing data

To avoid listwise deletion of individuals with missing data on confounding factors or outcomes, we performed multiple imputation of data assumed to be missing at random.[27] This was done using the fully conditional specification and sequential chained equations commands in Stata to create 100 imputed datasets. All variables in the analytical logistic regression models described above were included in separate imputation models for each relevant outcome. The pooling of RRs with 95% CIs across the 100 imputed datasets was done by applying Rubin's combination rules.[28]

### 2.6 | Ethics approval

The study was approved by the Norwegian Regional Committee for Medical and Health Research Ethics 2019/09/25 (10,145/2019/544). The Data Protection Impact Assessment (DPIA) was approved by the Data Protection Officer at the University of Bergen on 2020/08/26.

## 3 | RESULTS

We identified 55,386 eligible births divided into two cohorts of 7733 and 47,653 births dated by first- and second-trimester ultrasound respectively (Figure 1). The percentage of mothers older than 35 years at birth was higher in the first- than in the second-trimester cohort. They were also more often born in the Nordic countries, had more frequently completed higher education and had a lower prevalence of smoking compared to mothers in the second-trimester cohort. Other characteristics of the study population are listed in Table 1.

The ratios of male-to-female births by gestational week in different estimation methods are shown in Figure 2. The sex ratio in pregnancies dated by second-trimester ultrasound showed a pattern of male predominance in the early-term and the post-term period, and no clear sex predominance in births around weeks 38–40. The corresponding male-to-female sex ratio in pregnancies dated by first-trimester ultrasound showed less variation across gestational weeks. In pregnancies conceived by ART, we observed a similar pattern to that of the sex ratio from the first-trimester estimation method. The sex ratio showed less variation in iatrogenic than spontaneous births for all estimation methods (Data S2).

In the assessment of adverse outcomes for male infants born in the early-term period, we found reduced risk of NICU admission as well as of the composite of adverse outcomes in the first- compared to the second-trimester cohort (Table 2). Estimates of Apgar score $<7$ at 5 min and respiratory morbidity suggested a similar protective association in the first-trimester cohort. We found no reduction in risk of umbilical artery pH $<7.05$ in the first-trimester estimations, and the median Apgar score and mean umbilical artery pH were similar in the two cohorts.

For female infants born post-term in the first-trimester cohort, we found a reduction in risk estimates of umbilical artery pH $<7.05$, NICU admission, respiratory morbidity and the composite of adverse outcomes compared to the second-trimester cohort (Table 3). No difference in the likelihood of Apgar score $<7$ at 5 min, median Apgar score or mean umbilical artery pH was found between the two cohorts. We observed no differences in the risk of adverse outcomes in early-term females or post-term males between the cohorts (Data S3).

## 4 | COMMENT

### 4.1 | Principal findings

We found that the male-to-female sex ratio at birth in the early-term and post-term periods differed by first- compared to second-trimester ultrasound estimation of gestational age and that the first-trimester estimations showed a distribution of births by sex more similar to pregnancies conceived by ART. Furthermore, male infants born in gestational weeks 37–38 in pregnancies dated in the first trimester had a lower risk of NICU admission and a composite of adverse outcomes compared to those with second-trimester estimations. Females born at ≥41 weeks gestation had a lower risk of umbilical artery pH $<7.05$, NICU admission, respiratory morbidity and a composite of adverse outcomes in the earlier ultrasound prediction cohort; however, confidence intervals were wide.

**TABLE 1** Characteristics of 55,386 pregnancies by two cohorts of ultrasound estimation of gestational age, Medical Birth Registry of Norway, 2016–2020.

| | First-trimester estimation (n = 7733) n (%) | Second-trimester estimation (n = 47,653) n (%) |
|---|---|---|
| Foetal sex | | |
| Male | 3928 (50.8) | 24,448 (51.3) |
| Female | 3805 (49.2) | 23,205 (48.7) |
| Maternal age (years) | | |
| <25 | 462 (6.0) | 5755 (12.1) |
| 25–29 | 2155 (27.9) | 16,907 (35.5) |
| 30–34 | 2745 (35.5) | 17,322 (36.3) |
| ≥35 | 2371 (30.6) | 7669 (16.1) |
| Parity | | |
| 0 | 2947 (38.1) | 18,692 (39.2) |
| 1 | 2934 (37.9) | 17,972 (37.7) |
| 2 | 1336 (17.3) | 8028 (16.9) |
| ≥3 | 516 (6.7) | 2961 (6.2) |
| Maternal country of birth | | |
| Nordic countries | 6352 (84.5) | 34,536 (77.2) |
| Other | 1164 (15.5) | 10,200 (22.8) |
| Missing | 217 | 2917 |
| Marital status | | |
| Married/co-habitant | 7429 (96.1) | 45,647 (95.8) |
| Other | 301 (3.9) | 1984 (4.2) |
| Missing | 3 | 22 |
| Years of education | | |
| <10 | 786 (10.5) | 7059 (15.8) |
| ≥10 | 6730 (89.5) | 37,677 (84.2) |
| Missing | 217 | 2917 |
| Chronic maternal disease[a] | | |
| Yes | 695 (9.0) | 3140 (6.6) |
| No | 7038 (91.0) | 44,513 (93.4) |
| Smoking at beginning of pregnancy | | |
| Yes | 337 (5.2) | 3466 (7.9) |
| No | 6106 (94.8) | 40,421 (92.1) |
| Missing | 1290 | 3766 |
| Apgar score at 5 min | | |
| <7 | 126 (1.6) | 691 (1.5) |
| ≥7 | 7607 (98.4) | 46,960 (98.5) |
| Missing | 0 (0.0) | 2 (0.0) |
| Umbilical artery pH | | |
| <7.05 | 64 (1.4) | 364 (1.3) |
| ≥7.05 | 4480 (98.6) | 28,342 (98.7) |
| Missing | 3189 | 18,947 |

**TABLE 1** (Continued)

| | First-trimester estimation (n = 7733) n (%) | Second-trimester estimation (n = 47,653) n (%) |
|---|---|---|
| Neonatal intensive care unit admission | | |
| Yes | 525 (6.8) | 3809 (8.0) |
| No | 7208 (93.2) | 43,842 (92.0) |
| Missing | 0 | 2 |
| Respiratory morbidity[b] | | |
| Yes | 173 (2.2) | 1100 (2.3) |
| No | 7560 (97.8) | 46,553 (97.7) |

[a]Maternal asthma, pregestational diabetes, chronic kidney disease, hypertension, epilepsy or rheumatoid arthritis.

[b]Respiratory distress syndrome, treatment with continuous positive airway pressure (CPAP) or respirator.

### 4.2 | Strengths of the study

A main strength of our study was the reliability of the data source MBRN,[29] and the inclusion of a nearly complete population of births, which accounts for high external validity. Cohorts were sampled from the same period, and any observed differences in outcome between the cohorts are more likely a result of the ultrasound methods themselves than changes in clinical management policies. As no major structural changes in obstetric or neonatal care were introduced during the study period, adjustment for time variation was considered unnecessary.

### 4.3 | Limitations of the data

The first-trimester cohort comprised fewer births than the second-trimester cohort, which resulted in lower precision in the estimates. The limited sample in the first-trimester cohort made comparison of rare outcomes such as stillbirth and neonatal death unfeasible, and similarly hampered investigations of outcomes at preterm gestations. Also, differences between the cohorts in unmeasured factors affecting early foetal growth such as nutrition, infection and substance use could possibly bias ultrasound estimation of gestational age. Our study did not include data on maternal BMI, which could affect both frequencies of outcomes and validity of ultrasound measurements in both cohorts. Clinical indications for first-trimester ultrasound are independent risk factors for adverse neonatal outcomes, and although the issue was addressed by inclusion of covariates in the multivariable model, unmeasured or unknown risk factors could exist. Any such risk factor resulting in selection to the first-trimester estimation cohort could have attenuated the observed risk reduction in our study.

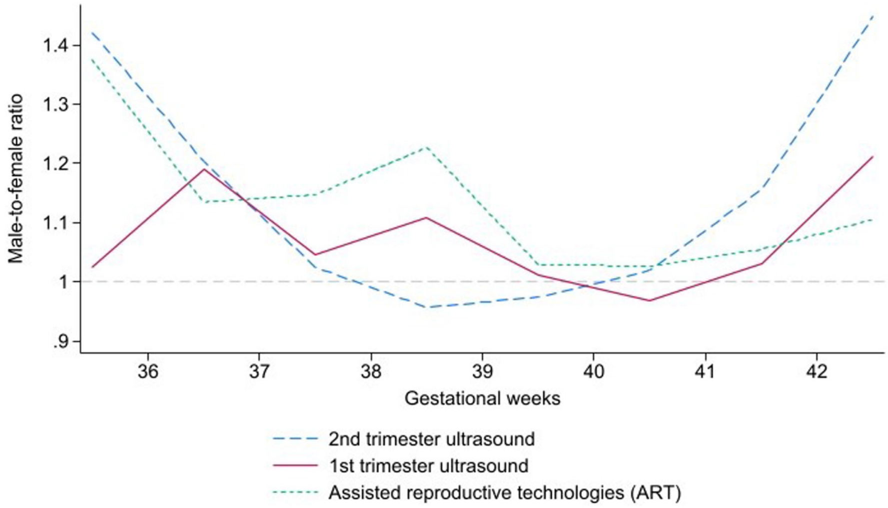


**FIGURE 2** Male-to-female ratio of infants born per gestational week in three cohorts where first-trimester ultrasound, second-trimester ultrasound and assisted reproductive technologies (ART) were used for estimation of gestational age.

**TABLE 2** Estimated relative risks (RR) with 95% confidence intervals (CI), medians with interquartile ranges (IQR) and means with standard deviations (SD) for adverse perinatal outcomes between two cohorts of ultrasound estimation method in male infants born in gestational week 37–38, Medical Birth Registry of Norway, 2016–2020.

| | First trimester *n* = 748 | Second trimester *n* = 3722 | | |
|---|---|---|---|---|
| | *n* (%) | *n* (%) | RR (95% CI) | aRR (95% CI)[a] |
| 5 min Apgar <7 | 7 (0.9) | 52 (1.4) | 0.67 (0.31, 1.47) | 0.63 (0.28, 1.41) |
| Umbilical artery pH <7.05 | 6 (0.8) | 23 (0.6) | 1.30 (0.53, 3.18) | 1.10 (0.44, 2.79) |
| NICU admission | 60 (8.0) | 379 (10.2) | 0.79 (0.61, 1.03) | 0.76 (0.58, 0.99) |
| Respiratory morbidity[b] | 13 (1.7) | 87 (2.3) | 0.74 (0.42, 1.33) | 0.68 (0.37, 1.25) |
| Composite score[c] | 66 (8.8) | 423 (11.4) | 0.77 (0.60, 0.98) | 0.74 (0.58, 0.95) |
| | | Median (IQR) | | Median (IQR) |
| Apgar score at 5 min | | 10 (1) | | 10 (1) |
| | | Mean (SD) | | Mean (SD) |
| Umbilical artery pH | | 7.25 (0.77) | | 7.26 (0.76) |

[a]Adjusted for maternal age, level of education, country of birth, smoking and chronic maternal illness.

[b]Respiratory distress syndrome, treatment with continuous positive airway pressure (CPAP) or respirator.

[c]Composite of adverse outcomes (5 min Apgar <7, umbilical artery pH <7.05, admission to neonatal intensive care unit (NICU) and respiratory morbidity).

### 4.4 | Interpretation

Kullinger et al.[11] demonstrated that Apgar score <7 at 5 min, hyperbilirubinemia, pneumothorax and RDS were less frequent in females born in weeks 37–38 compared to males at the same gestational age when estimated by ultrasound compared to LMP. These findings support the hypothesis of a sex-related bias in ultrasound estimation of gestational age, resulting in an increased early-term male risk of adverse outcomes at birth. As the study did not have information on gestational age at the time of the ultrasound examination, the authors hypothesized that any first-trimester estimation in this cohort would serve only to attenuate differences by reducing sex bias relative to second-trimester estimations. Findings from our study support this assumption since a direct comparison between the methods suggests that the excess risk of adverse outcomes in males born early term is reduced by first-trimester estimation, although not assessing the same prematurity-specific outcomes. The exception in our study was a point estimate of 1.10 for umbilical artery pH <7.05, however, the CI was wide. The proportions of missing values are comparable between first- and second-trimester estimated pregnancies indicating that differences in sampling are unlikely to explain this finding (Table 1). We found no difference in mean pH between the cohorts; thus, caution should be shown when interpreting this estimate.

For females born post-term, our study indicates that there might be a lower risk of adverse outcomes in the first- compared to the second-trimester cohort, supporting the hypothesis that early ultrasound results in less female underestimation of gestational age and consequently reduced excess post-term risk.[30] As less misclassification moves more female births to the post-term period, a reduction in male predominance is observed.

**TABLE 3** Estimated relative risks (RR) with 95% confidence intervals (CI), medians with interquartile ranges (IQR) and means with standard deviations (SD) for adverse perinatal outcomes between two cohorts of ultrasound estimation method in female infants born ≥41 weeks gestation, Medical Birth Registry of Norway, 2016–2020.

| | First trimester *n* = 884 | Second trimester *n* = 5522 | | |
|---|---|---|---|---|
| | *n* (%) | *n* (%) | RR (95% CI) | aRR (95% CI)[a] |
| 5 min Apgar <7 | 12 (1.4) | 72 (1.3) | 1.04 (0.57, 1.91) | 1.04 (0.55, 1.97) |
| Umbilical artery pH <7.05 | 9 (1.0) | 52 (0.9) | 1.08 (0.54, 2.19) | 0.93 (0.46, 1.87) |
| NICU admission | 36 (4.1) | 336 (6.1) | 0.67 (0.48, 0.94) | 0.71 (0.51, 1.00) |
| Respiratory morbidity[b] | 11 (1.2) | 90 (1.6) | 0.77 (0.41, 1.42) | 0.81 (0.42, 1,55) |
| Composite score[c] | 48 (5.4) | 400 (7.2) | 0.75 (0.56, 1.00) | 0.78 (0.58, 1.04) |
| | | Median (IQR) | | Median (IQR) |
| Apgar score at 5 min | | 10 (1) | | 10 (1) |
| | | Mean (SD) | | Mean (SD) |
| Umbilical artery pH | | 7.24 (0.08) | | 7.24 (0.08) |

[a]Adjusted for maternal age, level of education, country of birth, smoking and chronic maternal illness.

[b]Respiratory distress syndrome, treatment with continuous positive airway pressure (CPAP) or respirator.

[c]Composite of adverse outcomes (5 min Apgar <7, umbilical artery pH <7.05, admission to neonatal intensive care unit (NICU) and respiratory morbidity).

An increase in post-term labour inductions in Norway since 1999 has resulted in fewer severe post-term pregnancies as well as a reduction in total post-term birth proportions.[31] Consequently, the reported post-term male predominance was less pronounced in our study compared to previous reports.[10,13] As the risk of adverse outcomes increases with gestational age in the post-term period,[32] it is plausible that a policy of earlier induction of labour will attenuate the risk of sex-specific adverse outcomes from ultrasound estimation of gestational age.

In a secondary analysis of early-term females and post-term males, we found no or minor reductions in the risk of the same adverse neonatal outcomes in first- versus second-trimester ultrasound estimation (Data S3). Compared with the main analysis, these findings indicate that the effect on neonatal outcomes from reduced misclassification in early ultrasound varies across strata of sex and gestational age at birth. Finally, the comparison suggests that earlier ultrasound is more beneficial when it prevents moving births at higher risk gestation to lower risk gestation, namely preterm males classified as early-term and severely post-term females classified as post-term. Although the absolute risk reduction in 1st trimester ultrasound was modest in our study (Data S5), these cohorts represent a large proportion of women giving birth which implies a substantial impact on a population level.

Gestational age is arguably the most important measurement in reproductive epidemiology. Given its central importance, it is remarkable that gestational age can be estimated only approximately.[33] Birth at term in a healthy pregnancy requires limited medical support and surveillance. By contrast, preterm and post-term births have a higher risk of respiratory morbidity, cerebral palsy, neonatal infections and stillbirth.[34,35] Since differentiation of care is based on gestational age, misclassifying high-risk pregnancies to a low-risk group may lead to inadequate management and increased neonatal morbidity and mortality. Caughey et al. demonstrated that pregnancies dated in the first trimester had fewer post-term pregnancies beyond 41 and 42 weeks of gestation with a higher frequency of low 5 min Apgar score, haemorrhage, chorioamnionitis and macrosomia, but did not include sex-specific analyses.[36] The authors propose that accurate classification of gestational age redistributes post-term-specific risk to a more correct population. Our study suggests that this redistribution of risk is a result of two factors, namely fewer post-term males and more term females composing the post-term population in the first-trimester compared to second-trimester dating.

The sum of unknown and unmeasured factors affecting early foetal growth will bias any ultrasound prediction from a standard chart of gestational age, and their impact will vary by which foetal parameter is measured.[37] Detailed quantification of these factors and knowledge of their impact at different gestational ages will help determine whether any observed growth deviation is subject to a physiological or pathological process requiring clinical attention. Gestational age-specific estimates of neonatal risk will be biased by any exposure affecting growth, including foetal sex,[4] and the earlier a precise estimation of age can be obtained, the less biased the estimate will be. Of note, measurement error is also present in early ultrasound due to interobserver variation and impaired visualization, for instance, in obese patients. As females tend to be smaller than males also in the first trimester, errors in their measurements are potentially larger.

In our assessment of the male-to-female birth ratio by gestational week, we found the distribution of the sex ratio in ART pregnancies to be very similar to that of the first-trimester cohort (Figure 2). ART pregnancies represent a subgroup with an increased risk of several adverse pregnancy outcomes.[38,39] In addition to the observed difference in sex ratio between iatrogenic and spontaneous births that applied to all estimation methods, iatrogenic births were somewhat more frequent in the ART cohort (Data S2). The exact time of conception in spontaneous pregnancies can only be approximated, which makes evaluation of estimation models difficult, as there is no gold standard available for comparison. Although embryo transfers in ART provide a gold standard proxy for gestational age estimation, different obstetric management may bias a comparison with spontaneously conceived pregnancies.

We observed an increased male-to-female ratio in late preterm and early-term births in all estimation methods, contrary to the hypothesis of an estimation bias resulting in more female foetuses being born at earlier gestations. A similar finding was reported by Koch et al.,[13] who suggested that the increase may be the manifestation of an inherent preterm birth risk in male foetuses. Our study suggested that the late preterm male predominance was less pronounced in first-trimester ultrasound estimations, however, the sample size was small (Data S4). As the true unbiased sex ratio for births at different gestational ages remains unknown, caution is required when comparing the accuracy of different estimation methods. Although the sex ratios in ART and first-trimester ultrasound in our study were close to linear on a population level, it is biologically implausible that the sex ratio of births is identical at all gestational ages.

## 5 | CONCLUSIONS

The results of this population-based cohort study indicate less variation in sex distribution of births by gestational age and a reduction in early-term and post-term male predominance when ultrasound estimates are obtained earlier in pregnancy. Adverse neonatal outcomes were comparable or reduced in early-term males and post-term females in first-trimester compared to second-trimester estimations. Our findings support recommendations that first- is preferable to second-trimester ultrasound and highlight the importance of taking foetal sex and ultrasound estimation method into account when making clinical decisions based on gestational age.

### AUTHOR CONTRIBUTIONS

AE and NM conceived the idea, and LS and RN helped in designing the study. AE conducted all statistical analyses supervised by RN, produced all tables and figures and drafted the first version of the manuscript. CE provided specialist insight into obstetric ultrasound. AE and NM obtained funding for the study. All authors assisted with interpretation of results and revisions of the manuscript.

### ACKNOWLEDGEMENTS

We wish to thank Allen J. Wilcox for providing comments on an earlier version of the manuscript.

### FUNDING INFORMATION

The work was funded by the Western Norway Regional Health Authority (project number F-12156), Odd Fellow Medical fund and Gidske and Peter Jacob Sørensens fund.

### CONFLICT OF INTEREST STATEMENT

We declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

Data are available from the Medical Birth Registry of Norway and Statistics Norway, after a Norwegian Regional Committee for Medical and Health Ethics approval.

### CONSENT

No patient consent was obtained as data are available from the mandatory collection in the Medical Birth Registry of Norway and Statistics Norway. No material from other sources is reproduced in this manuscript.

### ORCID

*Anders Einum* https://orcid.org/0000-0002-2972-4868
*Cathrine Ebbing* https://orcid.org/0000-0002-4331-1250

### REFERENCES

1. Salomon LJ, Alfirevic Z, Berghella V, et al. ISUOG practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol*. 2022;59:840-856.
2. Kaelin Agten A, Xia J, Servante JA, Thornton JG, Jones NW. Routine ultrasound for fetal assessment before 24 weeks' gestation. *Cochrane Database Syst Rev*. 2021;8:CD014698.
3. Mongelli M, Wilcox M, Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. *Am J Obstet Gynecol*. 1996;174:278-281.
4. Henriksen TB, Wilcox AJ, Hedegaard M, Secher NJ. Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology*. 1995;6:533-537.
5. Vietheer A, Kiserud T, Haaland OA, Lie RT, Kessler J. Effect of maternal sleep on embryonic development. *Sci Rep*. 2022;12:17099.
6. Mook-Kanamori DO, Steegers EA, Eilers PH, Raat H, Hofman A, Jaddoe VW. Risk factors and outcomes associated with first-trimester fetal growth restriction. *JAMA*. 2010;303:527-534.
7. Workalemahu T, Grantz KL, Grewal J, Zhang C, Louis GMB, Tekola-Ayele F. Genetic and environmental influences on fetal growth vary during sensitive periods in pregnancy. *Sci Rep*. 2018;8:7274.
8. Divon MY, Ferber A, Nisell H, Westgren M. Male gender predisposes to prolongation of pregnancy. *Am J Obstet Gynecol*. 2002;187:1081-1083.
9. Galjaard S, Ameye L, Lees CC, et al. Sex differences in fetal growth and immediate birth outcomes in a low-risk Caucasian population. *Biol Sex Differ*. 2019;10:48.
10. Skalkidou A, Kieler H, Stephansson O, Roos N, Cnattingius S, Haglund B. Ultrasound pregnancy dating leads to biased perinatal morbidity and neonatal mortality among post-term-born girls. *Epidemiology*. 2010;21:791-796.
11. Kullinger M, Haglund B, Kieler H, Skalkidou A. Effects of ultrasound pregnancy dating on neonatal morbidity in late preterm and early term male infants: a register-based cohort study. *BMC Pregnancy Childbirth*. 2016;16:335.
12. 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;24:51-57.
13. Koch S, Lynggaard M, Jensen MS, Henriksen TB, Uldbjerg N. Sex bias in ultrasound measures of gestational age: assessment by sex ratio in post-term births. *Epidemiology*. 2014;25:513-517.

14. Irgens LM. The medical birth registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstet Gynecol Scand*. 2000;79:435-439.
15. Kessler J, Acharya G, Eggeboe T, Haugen G, von Brandis P. Ultralydundersøkelser i den alminnelige svangerskapsomsorgen. [ePub]2020 [26.01.2023]. Available from: https://www.legeforeningen.no/foreningsledd/fagmed/norsk-gynekologisk-forening/veiledere/veileder-i-fodselshjelp/ultralydundersokelser-i-den-alminnelige-svangerskapsomsorgen/
16. Morken NH, Haavaldsen CH, Holdø R, Øian B. P. Overtidig svangerskap. [ePub]2014 [cited 2023 20.02]. Available from: https://www.legeforeningen.no/foreningsledd/fagmed/norsk-gynekologisk-forening/veiledere/arkiv-utgatte-veiledere/veileder-i-fodselshjelp-2014/35.-overtidig-svangerskap/
17. Chen HY, Blackwell SC, Chauhan SP. Association between apgar score at 5 minutes and adverse outcomes among low-risk pregnancies. *J Matern Fetal Neonatal Med*. 2022;35:1344-1351.
18. Leinonen E, Gissler M, Haataja L, et al. Low Apgar scores at both one and five minutes are associated with long-term neurological morbidity. *Acta Paediatr*. 2018;107:942-951.
19. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med*. 2001;344:467-471.
20. Yli BM, Kessler J, Eikeland T, et al. Fosterovervåkning under fødsel, avnavling og syre-baseprøver fra navlesnor. [ePub]2014 [cited 2023 21.03]. Available from: https://www.legeforeningen.no/foreningsledd/fagmed/norsk-gynekologisk-forening/veiledere/arkiv-utgatte-veiledere/#Veileder%20i%20f%C3%B8dselshjelp%202020%20-%20Utg%C3%A5tte%20kapitler
21. Bovbjerg ML, Dissanayake MV, Cheyney M, Brown J, Snowden JM. Utility of the 5-minute Apgar score as a research endpoint. *Am J Epidemiol*. 2019;188:1695-1704.
22. Armstrong L, Stenson BJ. Use of umbilical cord blood gas analysis in the assessment of the newborn. *Arch Dis Child Fetal Neonatal Ed*. 2007;92:F430-F434.
23. Sitras V, Ulriksen M, Benth JS, Haugen G. Pregnant women's attitudes to prenatal screening in Norway. *Tidsskr nor Laegeforen*. 2020;140:1451-1455. https://pubmed.ncbi.nlm.nih.gov/33070591/
24. Matijasevich A, Victora CG, Lawlor DA, et al. Association of socioeconomic position with maternal pregnancy and infant health outcomes in birth cohort studies from Brazil and the UK. *J Epidemiol Community Health*. 2012;66:127-135.
25. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
26. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol*. 2016;45:1887-1894.
27. StataCorp. *Stata: Release 17: StataCorp LLC*. Stata Press; 2021.
28. Van Buuren S. *Flexible Imputation of Missing Data*. Chapman and Hall/CRC Press; 2018.
29. Moth FN, Sebastian TR, Horn J, Rich-Edwards J, Romundstad PR, Asvold BO. Validity of a selection of pregnancy complications in the medical birth registry of Norway. *Acta Obstet Gynecol Scand*. 2016;95:519-527.
30. Skalkidou A, Kullinger M, Georgakis MK, Kieler H, Kesmodel US. Systematic misclassification of gestational age by ultrasound biometry: implications for clinical practice and research methodology in the Nordic countries. *Acta Obstet Gynecol Scand*. 2018;97:440-444.
31. Haavaldsen C, Morken NH, Saugstad OD, Eskild A. Is the increasing prevalence of labor induction accompanied by changes in pregnancy outcomes? An observational study of all singleton births at gestational weeks 37-42 in Norway during 1999-2019. *Acta Obstet Gynecol Scand*. 2023;102:158-173.
32. Middleton P, Shepherd E, Crowther CA. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev*. 2018;5:CD004945.
33. Wilcox AJ. *Fertility and Pregnancy: An Epidemiologic Perspective*. 1st ed. Oxford University Press; 2010.
34. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet*. 2008;371:261-269.
35. Middleton P, Shepherd E, Morris J, Crowther CA, Gomersall JC. Induction of labour at or beyond 37 weeks' gestation. *Cochrane Database Syst Rev*. 2020;7:CD004945.
36. Caughey AB, Nicholson JM, Washington AE. First- vs second-trimester ultrasound: the effect on pregnancy dating and perinatal outcomes. *Am J Obstet Gynecol*. 2008;198:703.e1-703.e6. 703 e701-705; discussion 703 e705-706, 703.e1, 703.e6.
37. Salomon LJ. Early fetal growth: concepts and pitfalls. *Ultrasound Obstet Gynecol*. 2010;35:385-389.
38. Berntsen S, Soderstrom-Anttila V, Wennerholm UB, et al. The health of children conceived by ART: 'the chicken or the egg?'. *Hum Reprod Update*. 2019;25:137-158.
39. Westvik-Johari K, Romundstad LB, Lawlor DA, et al. Separating parental and treatment contributions to perinatal health after fresh and frozen embryo transfer in assisted reproduction: a cohort study with within-sibship analysis. *PLoS Med*. 2021;18:e1003683.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Einum A, Sørbye LM, Nilsen RM, Ebbing C, Morken N-H. Unveiling sex bias and adverse neonatal outcomes in ultrasound estimation of gestational age: A population-based cohort study. *Paediatr Perinat Epidemiol*. 2023;00:1-9. doi:10.1111/ppe.13029