Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hypertensive disorders of pregnancy after hormone replacement therapy versus natural cycle frozen embryo transfer: a propensity score-weighted cohort study

Goksu Goc

goksu.goc@gmail.com

American Hospital Kosovo https://orcid.org/0009-0003-1506-6061

Gülşah Aynaoğlu Yıldız

Ankara University https://orcid.org/0000-0002-3283-7783

Velsa Maloku

University of Prishtina

Tefta Isufaj Haliti

University of Prishtina https://orcid.org/0009-0001-9978-9253

Research Article

Keywords: Frozen embryo transfer, hormone replacement therapy, natural cycle, hypertensive disorders of pregnancy, pre-eclampsia, propensity score

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8437063/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

PREPRINT — Version 2. Not peer-reviewed. Supersedes version 1 (Research Square, DOI 10.21203/rs.3.rs-8437063/v1).

Hypertensive disorders of pregnancy after hormone replacement therapy versus natural cycle frozen embryo transfer: a propensity score-weighted cohort study

Goksu Goc[1]*, Gülşah Aynaoğlu Yildiz[2], Velsa Maloku[3], Tefta Isufaj Haliti[3]

[1]Department of Obstetrics and Gynecology, American Hospital Kosovo, Prishtina, Kosovo

[2]Department of Obstetrics and Gynecology, Faculty of Medicine, Ankara University, Ankara, Türkiye

[3]Clinic of Obstetrics and Gynecology, University Clinical Centre of Kosovo, Faculty of Medicine, University of Prishtina, Prishtina, Kosovo

***Correspondence:** Goksu Goc, MD, MSc; Department of Obstetrics and Gynecology, American Hospital Kosovo, Prishtina, Kosovo. E-mail: goksu.goc@gmail.com

ORCID: Goksu Goc 0009-0003-1506-6061; Gülşah Aynaoğlu Yildiz 0000-0002-3283-7783; Velsa Maloku; Tefta Isufaj Haliti 0009-0001-9978-9253

Version note: This version replaces version 1, which contained internal inconsistencies and two miscited references. The dataset was re-checked, the analysis was rewritten as a single reproducible script and independently audited, and every number was replaced. Appendix A lists the changes in seven groups: data file, analytical cohort, primary analysis, the perinatal death result, gestational-age units and growth definition, text, and the authorship,

Abstract

Background/aim: Hormone replacement therapy (HRT) for frozen embryo transfer (FET) suppresses the corpus luteum and has been linked to hypertensive disorders of pregnancy (HDP) in registries and randomised trials. We examined this association in a single-center cohort from Kosovo.

Materials and methods: This retrospective cohort included 2,943 autologous FET cycles (1,208 HRT, 1,735 natural cycle, including modified natural and mild stimulation cycles) at a single center, 2012–2024. Early pregnancy outcomes were analyzed per transfer; obstetric and neonatal outcomes among 1,341 singleton live births. The primary analysis used overlap weighting on a propensity score built from 20 variables. Risk ratios (RR) came from modified Poisson regression with cluster-robust standard errors. Growth categories used sex-specific INTERGROWTH-21st standards. Matching, restriction to 2016–2024, subtype analyses, and alternative specifications served as sensitivity analyses.

Results: HRT-FET had a lower live birth rate per transfer (43.4% vs 52.7%; RR 0.82, 95% CI 0.74–0.92). Among singleton live births, HRT-FET was associated with more HDP (12.2% vs 5.7%; RR 2.15, 95% CI 1.31–3.54; absolute risk difference +6.5 percentage points, 95% CI 2.0–11.1) and pre-eclampsia (8.5% vs 4.1%; RR 2.08, 95% CI 1.14–3.80). Matched-pair analysis gave HDP RR 3.86 (95% CI 1.68–8.88). The association persisted in 2016–2024 (RR 1.81, 95% CI

1.01–3.24). Birthweight was 89 g higher after HRT-FET (95% CI 12–166 g); small (RR 1.21, 95% CI 0.57–2.59) and large for gestational age (RR 1.25, 95% CI 0.86–1.81) and other outcomes showed no significant difference; the intervals do not establish equivalence. The E-value for HDP was 3.7.

Conclusion: HRT-FET was associated with a two-fold higher risk of HDP and a lower live birth rate per transfer than natural cycle FET. The direction agrees with recent randomised trials in ovulatory women. The findings support natural cycle preparation as the first option when ovulation can be monitored, and closer blood pressure surveillance with HRT-FET.

Key words: Frozen embryo transfer, hormone replacement therapy, natural cycle, hypertensive disorders of pregnancy, pre-eclampsia, propensity score

1. Introduction

Frozen embryo transfer (FET) has become a routine part of assisted reproduction, and freeze-all strategies have increased its use further [1,2]. The endometrium can be prepared in a natural cycle, with or without an ovulation trigger or mild stimulation, or with exogenous estrogen and progesterone (hormone replacement therapy, HRT) [3]. HRT cycles are easy to schedule and avoid cancellation for failed ovulation. They also suppress ovulation, so no corpus luteum forms.

The corpus luteum secretes relaxin, vascular endothelial growth factor, and other vasoactive factors that support maternal cardiovascular adaptation in early pregnancy [4–6]. Pregnancies without a corpus luteum show reduced aortic compliance and higher pre-eclampsia risk [7]. Registry and large cohort studies from Sweden, Japan, Denmark, France, and China, and several meta-analyses, report roughly two-fold higher odds of hypertensive disorders of pregnancy (HDP) after programmed FET than after natural cycle FET [8–15]. Findings for neonatal outcomes are less consistent [12,14,15].

Randomised trials have compared the two approaches mainly for live birth. The ANTARCTICA trial found modified natural and artificial cycles non-inferior for live birth [16]. Three recent trials in ovulatory women found similar live birth rates with natural, modified natural, and artificial preparation [17–19], while a fourth found a higher live birth rate with natural cycle preparation as the starting strategy [20]. The largest of these, with 4,376 women, also reported fewer hypertensive disorders and less pre-eclampsia after natural ovulation than after a programmed regimen [19]. A meta-analysis of observational studies reached the same conclusion for hypertensive disorders [21].

Registry data adjust for a limited set of covariates, and trials are powered for live birth rather than maternal complications. Independent cohorts with detailed treatment and embryo data, from settings other than the large registries, can test whether the association holds outside trial populations and can examine natural cycle subtypes separately. We therefore compared obstetric and neonatal outcomes after HRT-FET and natural cycle FET in a single-center cohort in Kosovo, using overlap weighting as the primary analysis, with matching, subtype, era-restricted, and negative control analyses as complements [22].

2. Materials and methods

2.1. Design, setting, and ethics

This retrospective cohort study included FET cycles performed at American Hospital Kosovo, Prishtina, between January 2012 and December 2024. The Kosovo Doctors Chamber Ethics Committee approved the study (reference 153/2025, 12 May 2025). Because the study is retrospective and uses de-identified data from medical records, the ethics committee waived the requirement for informed consent. The study followed the STROBE recommendations [23].

2.2. Participants

All autologous FET cycles with a known endometrial preparation protocol were eligible. Cycles using donor gametes and cycles with preimplantation genetic testing for monogenic disorders were not performed in the program. Cycles converted from a natural cycle to HRT because of failed follicular development (n = 57) were excluded. We analyzed obstetric and neonatal outcomes among singleton live births. Plurality was recorded only once, as the number of fetal hearts at the 6–7-week scan; gestational sac counts and later scans were not available in the extract. A singleton is therefore a live birth from a pregnancy with one fetal heart at that scan. Pregnancies with two fetal hearts were excluded as multiples even if only one infant was born, and a twin that vanished before the first scan cannot be identified. Data were extracted from the electronic medical record system of the hospital, and records with internal inconsistencies were verified against source documents before analysis (see Data verification).

2.3. Endometrial preparation

Natural cycle FET was offered to women with regular cycles. Monitoring started on cycle day 10–12. Ovulation was identified by a luteinizing hormone surge or follicle rupture (spontaneous natural cycle) or triggered with 5,000–10,000 IU human chorionic gonadotropin when the leading follicle reached 18–20 mm (modified natural cycle). Women with poor follicular development in a previous natural cycle received letrozole 2.5–5 mg or clomiphene citrate 50–100 mg daily on cycle days 3–7 (mild stimulation). All natural-type cycles received vaginal micronized progesterone 400–600 mg daily from the day of ovulation. These four subtypes formed the natural cycle group; all retain a corpus luteum.

HRT-FET was used for anovulation, irregular cycles, or scheduling needs. Oral estradiol valerate 4–8 mg daily or transdermal estradiol was started on cycle day 2–3, after pituitary downregulation with triptorelin 3.75 mg in about 40% of cycles. When endometrial thickness reached 7 mm or more, vaginal micronized progesterone 600–800 mg daily or intramuscular progesterone 50–100 mg daily was added. Cleavage-stage embryos were transferred after 3 days and blastocysts after 5 days of progesterone. Estrogen and progesterone continued until 10–12 weeks of gestation.

2.4. Outcomes

Per-cycle outcomes were positive serum β-human chorionic gonadotropin (10 IU/L or more at 12–14 days after transfer), clinical pregnancy (intrauterine sac with fetal heartbeat at 6–7 weeks), miscarriage (loss of a clinical pregnancy before 20 weeks, expressed per clinical pregnancy), and live birth. Positive tests without a clinical pregnancy (biochemical losses and ectopic pregnancies) were not counted as miscarriages. Cycles cancelled before transfer were not recorded, so live birth is per transfer, not per started cycle.

Maternal outcomes among singleton live births were HDP (new-onset hypertension of 140/90 mmHg or more after 20 weeks, including gestational hypertension, pre-eclampsia, eclampsia, and HELLP syndrome), pre-eclampsia, gestational diabetes (75-g oral glucose tolerance test at 24–28 weeks, IADPSG criteria), postpartum hemorrhage (institutional definition: estimated blood loss above 500 mL after vaginal and 1,000 mL after cesarean delivery), and cesarean delivery [24,25]. Neonatal outcomes were birthweight, gestational age at delivery, small (below the 10th centile) and large (above the 90th centile) for gestational age, low birthweight (below 2,500 g), macrosomia (above 4,000 g), preterm birth (before 37 weeks), neonatal intensive care unit (NICU) admission, and 5-min Apgar score. Congenital anomaly and neonatal death (death within the first 28 days after live birth, confirmed from the delivery and neonatal records; stillbirths are by definition absent from a live birth cohort) were analyzed as rare safety outcomes and used as negative control outcomes, on the assumption that the preparation protocol does not affect them directly; this assumption is weaker for neonatal death than for anomalies, and these results are exploratory [22]. Placenta previa, abruption, and uterine atony were documented mainly at cesarean delivery and were analyzed only in that subgroup. Diagnoses were taken from the delivery records as made by the treating obstetricians and were not re-adjudicated.

Gestational age at delivery was recorded as completed weeks and days (for example 37.6 for 37 weeks 6 days). It was converted to days for classification and to decimal weeks for continuous analysis. Small and large for gestational age were classified from birthweight, infant sex, and gestational age in days using the INTERGROWTH-21st Newborn Size Standards (birthweight below the sex-specific 10th centile or above the 90th centile), computed with the gigs package [26]. The standard covers 33 weeks 0 days to 42 weeks 6 days; one natural cycle birth at 43 weeks 2 days has no growth classification and remains in all other analyses, giving 1,340 births for growth outcomes. The clinical database also held SGA and LGA flags of unrecorded reference; these disagreed with the standard in 136 and 314 births respectively and are reported only as a sensitivity analysis. Preterm birth recalculated from days agreed with the recorded flag in every birth.

2.5. Covariates

The propensity score model included 20 variables: maternal age, body mass index, anti-Müllerian hormone, nulliparity, previous cesarean delivery, smoking status, prior miscarriage, previous HDP, gestational diabetes, postpartum hemorrhage and growth-restricted pregnancy, infertility factor, endometrial thickness, embryo stage, day of transfer, embryo quality (good, fair, poor), preimplantation genetic testing for aneuploidy, single versus double embryo transfer, transfer year, and number of previous transfers. These represented 26 model terms after dummy coding. Endometrial thickness was measured before progesterone was started; in HRT cycles this measurement follows estrogen exposure, so a sensitivity analysis omitted it. Luteal progesterone levels and other post-treatment variables were not included. No covariate had missing values.

2.6. Data verification

Before analysis, the analytic file was screened for internal inconsistencies (small and large for gestational age flagged together, preterm flag with recorded gestational age of 37 weeks or more, nulliparity with previous cesarean delivery, HRT cycles with endometrial thickness below 7 mm, and maternal age decreasing between cycles). The 243 flagged records were checked

against the source records by two clinical investigators; 268 values in 201 patients were corrected (maternal age 97, preterm flag 61, endometrial thickness 43, nulliparity 28, transfer year 24, SGA 10, LGA 5), infant sex was added, and each field-level correction was recorded in a change log identifying the record, the field, and the old and new values; the verifying clinician and source document for each entry are held by the authors. After verification, no logical inconsistency remained; two HRT cycles with endometrial thickness below 7 mm and six age sequences were confirmed as correct and retained. The analysis code halts if a logical inconsistency is present.

2.7. Statistical analysis

The propensity score was the predicted probability of HRT-FET from a logistic regression on the covariates above. The primary analysis used overlap weighting, which weights each woman by the probability of receiving the other protocol [27,28]. This targets the population in clinical equipoise, gives exact mean balance on every covariate, and does not rely on extreme weights. We checked balance using standardized mean differences (SMD), computed for binary and continuous variables alike as the difference in means divided by the pooled standard deviation, with values below 0.10 considered adequate. Risk ratios (RR) were estimated with weighted modified Poisson regression with cluster-robust standard errors by patient, because some women contributed more than one cycle [29]. Mean differences for continuous outcomes came from weighted linear regression with the same variance estimator; weighted means are shown with weighted standard deviations. Absolute risk differences were calculated from weighted risks and are the main measure of absolute effect; the number needed to harm is given for orientation only, because it assumes a causal effect. The E-value quantified the minimum strength of unmeasured confounding needed to explain the HDP association; it does not address selection, measurement error, or limited treatment overlap [30].

Sensitivity analyses followed a written analysis plan dated 27 October 2026, which distinguishes analyses planned in the first version of this study from those added after an independent methodological audit. Planned: 1:1 nearest-neighbor matching without replacement within a caliper of 0.2 standard deviations of the logit propensity score, with standard errors clustered by matched pair [31]; separate comparison of each natural cycle subtype with HRT-FET; exclusion of mild stimulation cycles; an exploratory analysis restricted to term births; and a test of interaction with single versus double embryo transfer. Added after the audit: restriction to 2016–2024, when both protocols were in routine use; era-stratified estimates with weights re-estimated within each era; a year-adjusted weighted model; the E-value; two alternative propensity model specifications (without endometrial thickness, and with transfer year in three categories); the INTERGROWTH-21st growth definition in place of the database flags; and, because 27 women appear in more than one matched pair, a matched analysis clustered by patient-linked groups of pairs instead of by pair. HDP and pre-eclampsia were the primary outcomes; no correction for multiple comparisons was applied to secondary outcomes, which are interpreted by effect size and consistency rather than by P values. Analyses used R 4.6.0 with MatchIt 4.7.2 [32], WeightIt 1.7.0, cobalt 4.6.3, sandwich 3.1.2 [33], and gigs 0.5.2; figures were drawn with Python 3.13 and matplotlib from the exported analytic data and tables. An independent computational re-analysis in Python reproduced the matched-pair memberships and the primary, matched, and era-restricted estimates. The analysis code is available from the corresponding author.

3. Results

3.1. Cohort

After exclusion of 57 converted cycles, 2,943 FET cycles from 1,865 women were analyzed: 1,208 HRT-FET (41.0%) and 1,735 natural cycle FET (431 spontaneous natural, 661 modified natural, 318 letrozole, 325 clomiphene). Protocol use changed sharply over time. Between 97% and 100% of cycles in 2012–2015 were HRT-FET, whereas HRT-FET made up 10–19% of cycles each year from 2016 to 2024 (Supplementary Table S1). Live birth occurred in 1,439 cycles; 98 were multiple births, leaving 1,341 singleton live births (489 HRT-FET, 852 natural cycle) for the obstetric analysis (Figure 1).

Before weighting, women in the HRT-FET group were younger, had higher anti-Müllerian hormone, fewer previous cesarean deliveries and growth-restricted pregnancies, and more double embryo transfers, and their cycles were earlier (Table 1). The propensity score model discriminated strongly (c-statistic 0.90), driven mainly by transfer year (SMD 1.77); age (0.35), anti-Müllerian hormone (0.28), single embryo transfer (0.27), previous cesarean delivery (0.21), previous growth restriction (0.18), prior miscarriage (0.14), tubal factor (0.14), current smoking (0.13), and body mass index (0.12) also exceeded 0.10. After overlap weighting the means of all model covariates were balanced to numerical precision (all SMD < 0.001; Table 1; Figure 2). The effective sample size was 267 HRT-FET and 404 natural cycle births.

3.2. Early pregnancy outcomes

In the weighted per-cycle analysis, HRT-FET had lower rates of positive β-hCG (56.4% vs 64.0%; RR 0.88, 95% CI 0.81–0.96), clinical pregnancy (53.0% vs 60.5%; RR 0.88, 95% CI 0.80–0.96), and live birth (43.4% vs 52.7%; RR 0.82, 95% CI 0.74–0.92; absolute difference −9.3 percentage points, 95% CI −14.3 to −4.3). Miscarriage per clinical pregnancy was more frequent after HRT-FET (18.1% vs 12.9%; RR 1.41, 95% CI 1.03–1.93). The singleton proportion among live births was similar (Table 2).

3.3. Maternal outcomes

HDP occurred in 51 of 489 HRT-FET and 63 of 852 natural cycle singleton births. In the overlap-weighted analysis the risk was 12.2% vs 5.7% (RR 2.15, 95% CI 1.31–3.54, P = 0.003), an absolute increase of 6.5 percentage points (95% CI 2.0–11.1), which corresponds to one additional case per 15 HRT-FET births if the association is causal. Pre-eclampsia was also more frequent (8.5% vs 4.1%; RR 2.08, 95% CI 1.14–3.80, P = 0.018; absolute increase 4.4 percentage points, 95% CI 0.4–8.4). The crude RRs were 1.41 (95% CI 0.99–2.01) for HDP and 1.20 (0.77–1.87) for pre-eclampsia, so adjustment moved the estimates away from the null. The E-value for the HDP point estimate was 3.7 and for the lower confidence limit 1.9. Gestational diabetes (RR 1.02), postpartum hemorrhage (RR 1.24), and cesarean delivery (RR 0.95) showed no significant difference (Table 3; Figure 3).

3.4. Neonatal outcomes

Mean birthweight was 89 g higher after HRT-FET (95% CI 12–166 g, P = 0.024); gestational age at delivery was 39.4 vs 39.2 weeks (difference +0.18 weeks, 95% CI −0.07 to 0.43). With the INTERGROWTH-21st definition, small for gestational age occurred in 19 of 489 and 34 of 851 births (weighted 4.7% vs 3.8%; RR 1.21, 95% CI 0.57–2.59) and large for gestational age in 89 and 117 (15.3% vs 12.2%; RR 1.25, 95% CI 0.86–1.81). With the database

flags instead, small for gestational age had appeared less frequent after HRT-FET (RR 0.47, 95% CI 0.25–0.87; Supplementary Table S9), a result that does not survive the standardized definition. Macrosomia, low birthweight, preterm birth, NICU admission, and 5-min Apgar score showed no significant difference; the upper confidence limits (2.18 for NICU admission, 2.00 for low birthweight, 1.81 for large for gestational age) do not exclude moderate increases. The rare outcomes were imprecise: congenital anomaly RR 1.63 (95% CI 0.53–4.99; 28 events) and neonatal death RR 0.18 (95% CI 0.03–1.10; 7 events) (Table 3). The first version of this analysis reported a significant inverse association with perinatal death (RR 0.12, P = 0.004); that analysis used all 3,000 cycles as the denominator and included eight death flags in cycles without a live birth that were corrected in the present dataset. Among singleton live births the death counts are the same in both datasets (2 vs 5).

3.5. Sensitivity analyses

One-to-one matching retained 188 pairs (38.4% of HRT-FET births). Residual imbalance remained for ovulatory factor (SMD 0.22), transfer year (0.18), endometriosis (0.18), number of previous transfers (0.16), previous growth restriction (0.16), single embryo transfer (0.12), and previous HDP (0.10), because few natural cycle births were available in the early years. The matched estimates were larger than the weighted ones but in the same direction: HDP 14.4% vs 3.7% (RR 3.86, 95% CI 1.68–8.88, P = 0.002) and pre-eclampsia 10.1% vs 3.2% (RR 3.17, 95% CI 1.26–7.95, P = 0.014). Birthweight was 101 g higher (95% CI 1–201 g); other neonatal outcomes showed no significant difference (small for gestational age RR 1.39, 95% CI 0.44–4.39; large for gestational age RR 1.25, 0.74–2.12). The 376 matched births came from 342 women; clustering by the 154 patient-linked groups of pairs instead of by pair left the estimates essentially unchanged (HDP RR 3.86, 95% CI 1.72–8.66) (Supplementary Table S2).

In matched comparisons of each natural cycle subtype with HRT-FET, the HDP risk ratio was 2.67 (95% CI 1.04–6.84) for spontaneous natural cycles (123 pairs), 2.50 (1.10–5.69) for modified natural cycles (149 pairs), 1.67 (0.60–4.61) for letrozole cycles (97 pairs), and 1.87 (0.79–4.44) for clomiphene cycles (112 pairs) (Supplementary Table S3). After exclusion of mild stimulation cycles (176 pairs), the risk ratio was 2.20 (95% CI 1.06–4.55) for HDP and 1.78 (0.78–4.03) for pre-eclampsia (Supplementary Table S4).

Restricted to 2016–2024 (956 singleton births, 109 HRT-FET), when protocol choice no longer tracked calendar time (c-statistic 0.67), HRT-FET remained associated with HDP (12.1% vs 6.7%; RR 1.81, 95% CI 1.01–3.24, P = 0.045) and pre-eclampsia (10.1% vs 5.0%; RR 2.02, 95% CI 1.05–3.87, P = 0.035). Live birth per transfer was also lower in this period (RR 0.82, 95% CI 0.70–0.95). Era-stratified estimates, with weights re-estimated within each era, were heterogeneous and imprecise: RR 4.92 (95% CI 1.60–15.1) for 2012–2017, where only 8 natural cycle births had HDP, and RR 1.71 (0.88–3.33) for 2018–2024. Adding transfer year to the weighted outcome model gave RR 2.11 (95% CI 1.30–3.40). Omitting endometrial thickness from the propensity model gave RR 2.18 (1.33–3.57), and entering transfer year as three categories gave RR 1.90 (1.08–3.35) (Supplementary Table S5).

Among term births (177 matched pairs), HDP remained more frequent after HRT-FET (RR 4.00, 95% CI 1.63–9.81); small for gestational age (RR 2.50, 95% CI 0.48–12.95), large for gestational age (RR 1.45, 0.85–2.48), and birthweight (+83 g, 95% CI −16 to +183 g) did not

differ significantly (Supplementary Table S6). Placental complications were documented in only 265 of 619 cesarean deliveries (57% missing); placenta previa occurred in 3 of 93 HRT-FET and 1 of 172 natural cycle deliveries, and no estimate from this subgroup should be interpreted (Supplementary Table S7). No interaction with single versus double embryo transfer reached significance (P for interaction 0.83 for HDP, 0.92 for pre-eclampsia, 0.87 for preterm birth, 0.08 for small for gestational age; Supplementary Table S8).

Table 1. Baseline characteristics of singleton live births by endometrial preparation protocol, before and after overlap weighting.

Characteristic	Natural cycle FET (n = 852)	HRT-FET (n = 489)	SMD before	Natural cycle, weighted	HRT-FET, weighted
Maternal age, years	34.6 (4.4)	33.1 (4.1)	0.35	33.8	33.8
Body mass index, kg/m²	24.1 (3.1)	23.7 (2.9)	0.12	23.8	23.8
Anti-Müllerian hormone, ng/mL	3.1 (1.0)	3.3 (0.9)	0.28	3.2	3.2
Nulliparous	303 (35.6%)	185 (37.8%)	0.05	32.4%	32.4%
Previous cesarean delivery	250 (29.3%)	100 (20.4%)	0.21	27.0%	27.0%
Prior miscarriage	168 (19.7%)	70 (14.3%)	0.14	15.4%	15.4%
Previous hypertensive disorder	70 (8.2%)	28 (5.7%)	0.10	7.0%	7.0%
Previous gestational diabetes	81 (9.5%)	38 (7.8%)	0.06	8.6%	8.6%
Previous postpartum hemorrhage	55 (6.5%)	24 (4.9%)	0.07	6.5%	6.5%
Previous growth restriction	85 (10.0%)	26 (5.3%)	0.18	7.6%	7.6%
Smoking status					
Never	630 (73.9%)	380 (77.7%)	0.09	75.7%	75.7%
Former	149 (17.5%)	83 (17.0%)	0.01	17.5%	17.5%
Current	73 (8.6%)	26 (5.3%)	0.13	6.8%	6.8%
Infertility factor					
Male factor	271 (31.8%)	152 (31.1%)	0.02	30.8%	30.8%
Unexplained	176 (20.7%)	99 (20.2%)	0.01	20.1%	20.1%
Ovulatory	141 (16.5%)	73 (14.9%)	0.04	19.2%	19.2%
Tubal	100 (11.7%)	82 (16.8%)	0.14	13.7%	13.7%
Endometriosis	113 (13.3%)	51 (10.4%)	0.09	10.2%	10.2%
Mixed	51 (6.0%)	32 (6.5%)	0.02	6.0%	6.0%
Endometrial thickness, mm	9.6 (1.8)	9.8 (1.4)	0.09	9.8	9.8
Embryo stage					
Cleavage stage	75 (8.8%)	37 (7.6%)	—	8.0%	8.0%
Blastocyst	777 (91.2%)	452 (92.4%)	0.05	92.0%	92.0%
Day of transfer	4.9 (0.7)	5.0 (0.7)	0.07	5.0	5.0
Embryo quality					
Good	569 (66.8%)	320 (65.4%)	0.03	67.9%	67.9%
Fair	153 (18.0%)	86 (17.6%)	0.01	16.8%	16.8%
Poor	130 (15.3%)	83 (17.0%)	0.05	15.3%	15.3%
PGT-A	173 (20.3%)	87 (17.8%)	0.06	19.5%	19.5%
Single embryo transfer	717 (84.2%)	358 (73.2%)	0.27	77.7%	77.7%
Transfer year	2019.8 (2.6)	2014.9 (2.9)	1.77	2017.5	2017.5
Previous embryo transfers, n	1.9 (1.0)	1.8 (1.0)	0.07	1.9	1.9

Values are mean (SD) or n (%); weighted columns give overlap-weighted means or percentages. SMD, absolute standardized mean difference before weighting, computed for binary and continuous variables as the difference in means divided by the pooled standard deviation; all SMD after overlap weighting were < 0.001. After 1:1 matching, SMD exceeded 0.10 for ovulatory factor (0.22), transfer year (0.18),

endometriosis (0.18), previous transfers (0.16), previous growth restriction (0.16), single embryo transfer (0.12) and previous hypertensive disorder (0.10). FET, frozen embryo transfer; HRT, hormone replacement therapy; PGT-A, preimplantation genetic testing for aneuploidy.

Table 2. Early pregnancy outcomes per transfer cycle, HRT-FET versus natural cycle FET, overlap-weighted analysis of 2,943 cycles.

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Positive β-hCG	686/1208 (56.4)	1105/1735 (64.0)	0.88 (0.81–0.96)	0.003	−7.6 (−12.5 to −2.7)
Clinical pregnancy	645/1208 (53.0)	1049/1735 (60.5)	0.88 (0.80–0.96)	0.004	−7.5 (−12.4 to −2.5)
Miscarriage (per clinical pregnancy)	113/645 (18.1)	142/1049 (12.9)	1.41 (1.03–1.93)	0.032	+5.3 (+0.3 to +10.2)
Live birth	532/1208 (43.4)	907/1735 (52.7)	0.82 (0.74–0.92)	<0.001	−9.3 (−14.3 to −4.3)
Singleton (per live birth)	489/532 (93.1)	852/907 (92.5)	1.01 (0.97–1.05)	0.752	+0.6 (−3.0 to +4.1)

Percentages are overlap-weighted. Live birth is per transfer; cancelled cycles were not recorded. Miscarriage is loss of a clinical pregnancy before 20 weeks, per clinical pregnancy. RR, risk ratio from modified Poisson regression with cluster-robust standard errors by patient; CI, confidence interval; pp, percentage points; hCG, human chorionic gonadotropin.

Table 3. Maternal and neonatal outcomes among 1,341 singleton live births, HRT-FET versus natural cycle FET: overlap-weighted primary analysis and 1:1 matched confirmatory analysis.

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR or MD (95% CI), weighted	P	Risk difference, pp (95% CI)	RR or MD (95% CI), matched (188 pairs)
Maternal outcomes						
Hypertensive disorders of pregnancy	51/489 (12.2)	63/852 (5.7)	2.15 (1.31–3.54)	0.003	+6.5 (+2.0 to +11.1)	3.86 (1.68–8.88)
Pre-eclampsia	31/489 (8.5)	45/852 (4.1)	2.08 (1.14–3.80)	0.018	+4.4 (+0.4 to +8.4)	3.17 (1.26–7.95)
Gestational diabetes	38/489 (9.6)	89/852 (9.4)	1.02 (0.62–1.68)	0.943	+0.2 (−4.6 to +4.9)	0.89 (0.47–1.69)
Postpartum hemorrhage	27/489 (6.7)	49/852 (5.4)	1.24 (0.67–2.31)	0.500	+1.3 (−2.6 to +5.2)	1.30 (0.57–2.97)
Cesarean delivery	209/489 (45.3)	410/852 (47.6)	0.95 (0.81–1.12)	0.562	−2.3 (−10.0 to +5.4)	0.98 (0.78–1.23)
Neonatal outcomes						
Small for gestational age	19/489 (4.7)	34/851 (3.8)	1.21 (0.57–2.59)	0.622	+0.8 (−2.5 to +4.1)	1.39 (0.44–4.39)
Large for gestational age	89/489 (15.3)	117/851 (12.2)	1.25 (0.86–1.81)	0.247	+3.0 (−2.2 to +8.2)	1.25 (0.74–2.12)
Low birthweight (<2,500 g)	18/489 (3.6)	33/852 (4.0)	0.90 (0.41–2.00)	0.804	−0.4 (−3.4 to +2.6)	0.71 (0.23–2.26)
Macrosomia (>4,000 g)	68/489 (11.6)	85/852 (8.9)	1.30 (0.84–2.01)	0.241	+2.7 (−1.9 to +7.2)	1.37 (0.73–2.58)

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR or MD (95% CI), weighted	P	Risk difference, pp (95% CI)	RR or MD (95% CI), matched (188 pairs)
Preterm birth (<37 weeks)	39/489 (7.0)	69/852 (7.9)	0.88 (0.52–1.51)	0.652	−0.9 (−4.9 to +3.0)	0.80 (0.41–1.57)
NICU admission	42/489 (9.7)	51/852 (7.5)	1.30 (0.78–2.18)	0.318	+2.2 (−2.3 to +6.8)	1.38 (0.69–2.77)
Birthweight, g	3398 ± 510	3309 ± 498	+89 (+12 to +166)	0.024	—	+101 (+1 to +201)
Gestational age at delivery, weeks	39.4 ± 1.6	39.2 ± 1.7	+0.18 (−0.07 to +0.43)	0.152	—	+0.24 (−0.09 to +0.57)
Apgar score at 5 min	9.6 ± 0.7	9.5 ± 0.7	+0.02 (−0.09 to +0.13)	0.715	—	+0.02 (−0.11 to +0.15)
Rare outcomes used as negative controls						
Congenital anomaly	10/489 (2.6)	18/852 (1.6)	1.63 (0.53–4.99)	0.394	+1.0 (−1.4 to +3.4)	2.00 (0.36–10.97)
Neonatal death	2/489 (0.2)	5/852 (1.1)	0.18 (0.03–1.10)	0.064	−0.9 (−2.2 to +0.4)	NE

Counts and unweighted denominators are shown; percentages are overlap-weighted. Continuous outcomes are weighted mean ± weighted SD in the weighted columns and mean ± SD in the matched column, with mean difference (MD). Small and large for gestational age use the sex-specific INTERGROWTH-21st definition in 1,340 births (489 HRT-FET, 851 natural cycle; one birth beyond 42 weeks 6 days excluded). Gestational age was recorded as weeks and days and converted to decimal weeks. Neonatal death is death recorded among live births. RR, risk ratio; CI, confidence interval; pp, percentage points; NICU, neonatal intensive care unit; NE, not estimable (no events in one group).

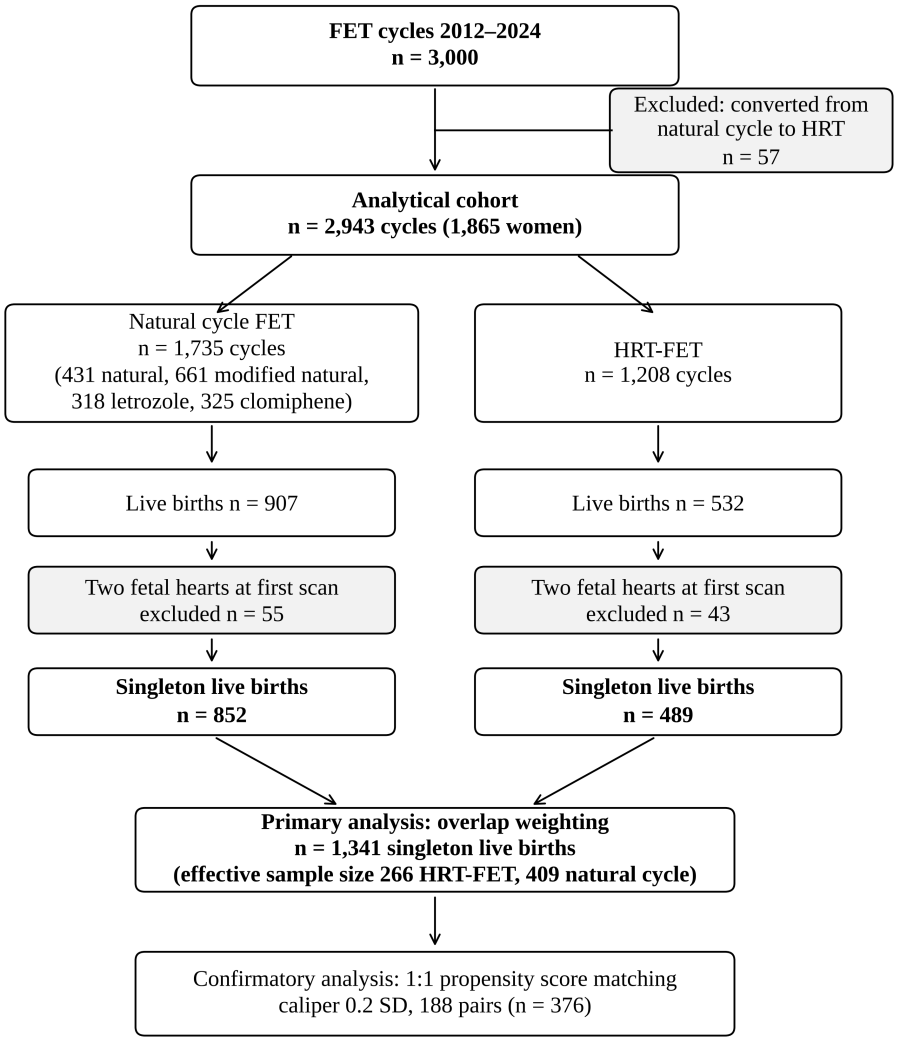


Figure 1. Flow of cycles through the study. HRT, hormone replacement therapy; FET, frozen embryo transfer.

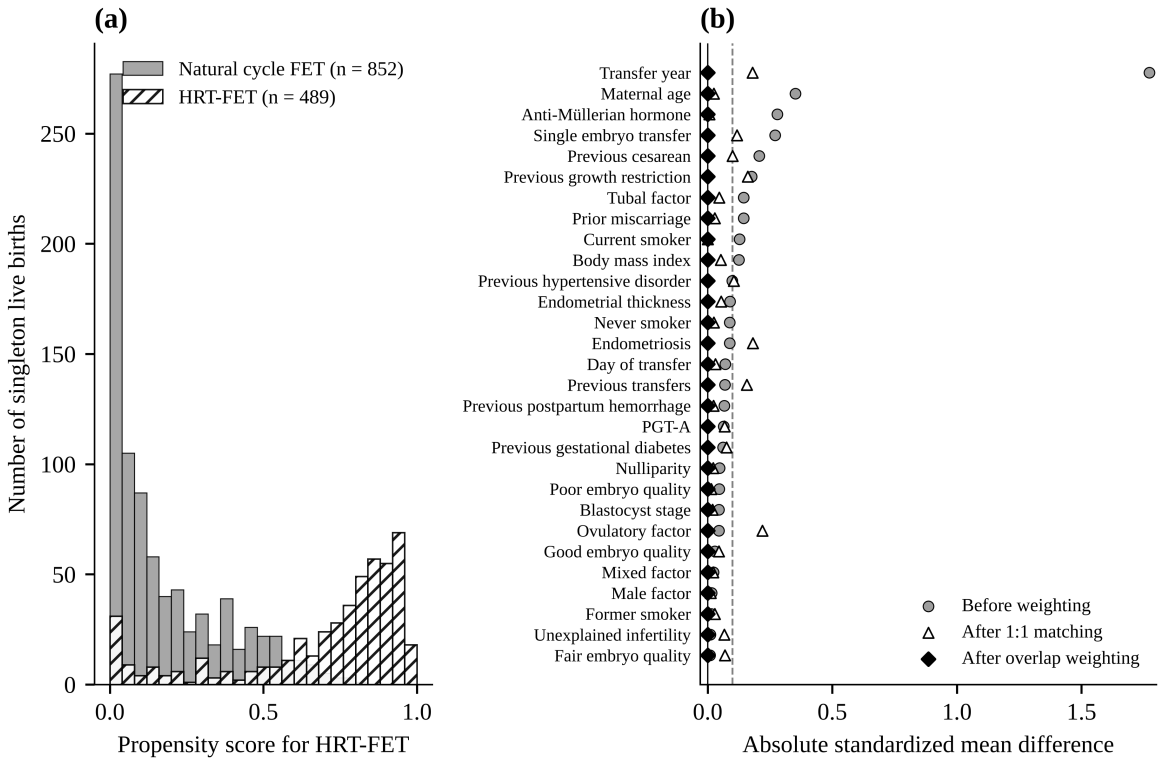


Figure 2. Propensity score diagnostics in singleton live births. (a) Propensity score distributions by protocol before weighting. (b) Absolute standardized mean differences before weighting (grey circles), after 1:1 matching (open triangles), and after overlap weighting (black diamonds), all computed from the analytic data and weights; the dashed line marks 0.10. Standardized mean differences use the pooled standard deviation for binary and continuous variables.

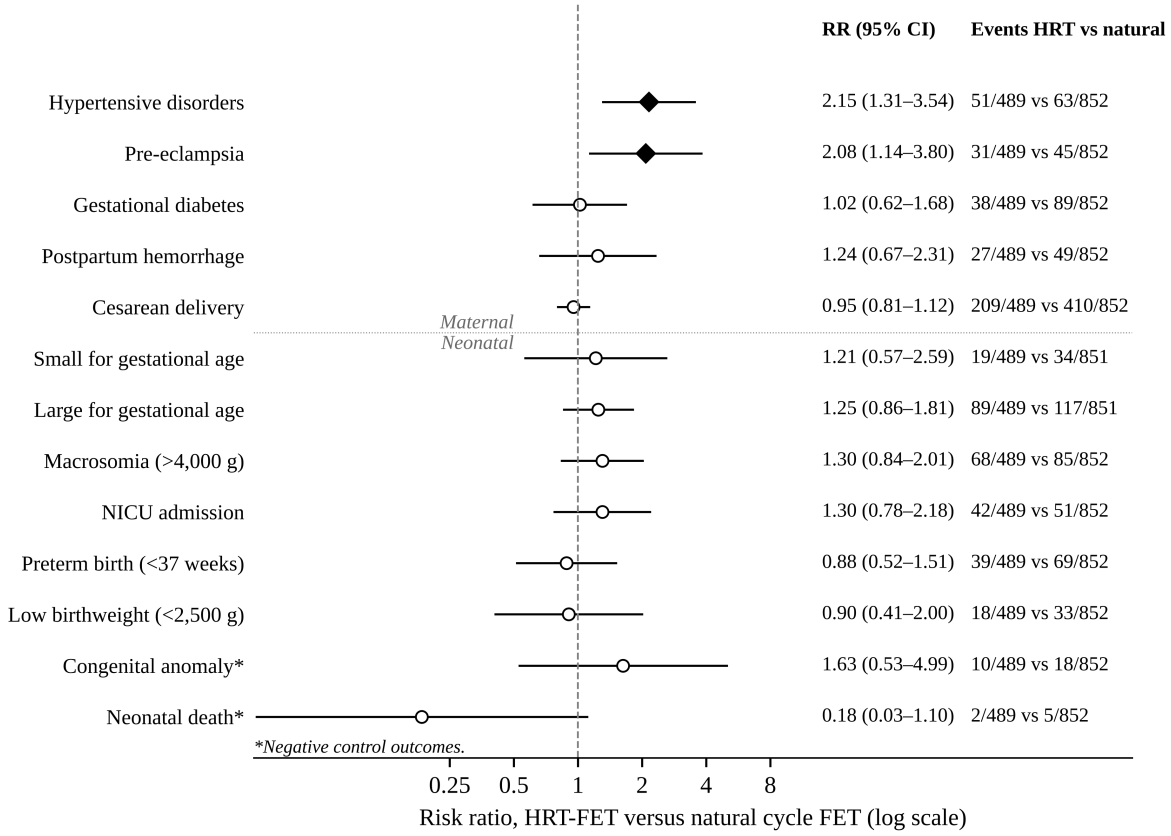


Figure 3. Risk ratios (95% confidence intervals) for maternal and neonatal outcomes, HRT-FET versus natural cycle FET, from the overlap-weighted analysis of 1,341 singleton live births (1,340 for small and large for gestational age, INTERGROWTH-21st definition). Values above 1 indicate higher risk with HRT-FET.

4. Discussion

4.1. Main findings

In this cohort of 2,943 FET cycles, HRT-FET was associated with an 18% lower live birth rate per transfer and, among singleton live births, a two-fold higher risk of HDP and pre-eclampsia than natural cycle FET. The absolute excess was about six HDP cases per 100 births. The association was present with overlap weighting and with matching, in the same direction within each natural cycle subtype, after exclusion of stimulated cycles, under alternative propensity model specifications, and in the years when both protocols were used side by side. We found no significant difference in prematurity, NICU admission, or large for gestational age, but the study was not powered to exclude moderate differences. Birthweight was slightly higher after HRT-FET, but small and large for gestational age, classified with a sex-specific international standard, showed no significant difference; the growth intervals are wide and do not establish equivalence.

4.2. Comparison with previous studies

Our relative estimate of 2.2 is close to the pooled odds ratios of 1.9 for HDP and 2.1 for pre-eclampsia in the meta-analysis of Busnelli et al., which rated the certainty of that evidence as very low and low [11], to the meta-analysis of Zaat et al. [21], and to registry data from Sweden [8], Japan [9], Denmark [10], and France [13]; the Moreno-Sepulveda meta-analysis reports the same contrast in the opposite direction, as risk ratios favouring natural cycles [12]. The first version of this analysis, based on a preliminary extract, reported a larger effect; the corrected dataset gives a more modest estimate. Most importantly, the direction agrees with the PnROVE trial, in which 4,376 ovulatory women were randomised and pre-eclampsia occurred in 2.9% versus 4.6% and HDP in 6.8% versus 10.0% of clinical pregnancies with natural versus programmed preparation [19]. Our cohort therefore adds an independent, real-world confirmation from a setting without registry coverage, rather than new evidence of a kind not previously available.

The live birth results are less consistent across studies. The Vietnamese and Belgian trials and PnROVE found similar live birth or clinical pregnancy rates with natural and artificial preparation [17–19], the COMPETE trial found a higher live birth rate with a natural cycle starting strategy (54% vs 43%) [20], and ANTARCTICA found modified natural cycles non-inferior to artificial cycles [16]. Our per-transfer difference of nine percentage points is larger than most of these and probably reflects confounding by indication that weighting cannot fully remove: in our program HRT-FET was often chosen for women with irregular cycles or ovulatory dysfunction, and before 2016 it was used for nearly everyone. Because cancelled cycles were not recorded, our live birth estimate is per transfer, which favours natural cycles relative to a per-started-cycle comparison in which cancellations count against them.

The slightly higher birthweight after HRT-FET is in the direction of Swedish and Japanese registry data showing more large for gestational age infants after programmed cycles [8,9], but our large for gestational age estimate was imprecise. An earlier draft of this analysis, using growth flags from the clinical database, had found fewer small for gestational age infants after HRT-FET; recalculation with sex-specific INTERGROWTH-21st standards removed that association, and the database flags disagreed with the standard in about a third of births. We

therefore report growth outcomes only under the standardized definition and draw no conclusion from the earlier flags. All neonatal analyses are conditioned on live birth, and selective early loss of HRT pregnancies with poor placentation, consistent with the higher miscarriage rate we observed, could bias growth comparisons among survivors in either direction [34].

4.3. Biological plausibility

The corpus luteum hypothesis offers a coherent mechanism. Relaxin and other luteal factors drive the fall in systemic vascular resistance and the rise in arterial compliance of early pregnancy [4,5]. Women conceiving without a corpus luteum show blunted vascular adaptation and more pre-eclampsia [7]. Angiogenic imbalance is central to pre-eclampsia [35], and placental epigenetic programming is sensitive to the periconceptional environment [36], but we did not measure hormone levels or angiogenic markers, so these remain hypotheses in our cohort. The similar direction of our estimates across natural cycle subtypes, each of which retains a corpus luteum, is compatible with this model but does not test it. Differences in progesterone route and dose, and GnRH agonist downregulation in some HRT cycles, are alternative explanations we could not separate [37].

4.4. Strengths and limitations

Strengths include detailed embryo and treatment data, a written analysis plan with several estimands, record-level data verification, and transparent reporting of the calendar-time structure of the cohort. Several limitations need emphasis. First, HRT-FET was essentially the only protocol before 2016, so comparisons across the whole study period lean on the propensity model to bridge calendar time. Overlap weighting gives most weight to women who could plausibly have received either protocol, so the estimate applies to that population and not to all FET patients; it also down-weights, but does not remove, the early HRT-only years. We therefore repeated the analysis in 2016–2024, where the association remained, and we regard that period as the most clinically comparable. The much larger and imprecise estimate in 2012–2017 rests on eight natural cycle events and may also reflect changes in blood pressure surveillance or aspirin use that we could not measure. Second, this is a single-center study with diagnoses taken from clinical records without re-adjudication. Misclassification may have occurred; its direction and effect on the estimates are uncertain. Third, obstetric outcomes were conditioned on singleton live birth, and plurality was known only from the fetal heart count at the first scan; we could not identify vanishing twins before that scan, and we excluded pregnancies that later lost one of two fetal hearts rather than analyzing them as singletons. Fourth, the reasons for choosing HRT (anovulation, irregular cycles, scheduling), polycystic ovary phenotype, chronic hypertension, pre-pregnancy diabetes, and aspirin prophylaxis were not recorded; the infertility factor variable captures only part of this. The E-value of 3.7 indicates that an unmeasured confounder would need to be associated with both protocol choice and HDP by a risk ratio of about 3.7 to explain the association; it says nothing about selection or measurement error. Fifth, the rare outcomes used as negative controls were imprecise (28 anomalies and 7 neonatal deaths) and cannot exclude residual bias; the change in the perinatal death result between versions of this analysis, explained in the Results, is a reminder that denominators matter as much as data. Sixth, several growth and prematurity flags in the clinical database were internally inconsistent; we corrected prematurity and parity against source records and reclassified growth from birthweight, sex, and gestational

age using a published standard rather than relying on the database flags. Residual misclassification of outcomes recorded by treating clinicians over 13 years remains possible and its direction is unknown. Finally, cycles cancelled before transfer were not recorded.

4.5. Clinical implications

Taken together with the randomised evidence, our data support natural cycle preparation, including modified natural cycles, as the first option for ovulatory women, because it carries a lower risk of HDP without a lower live birth rate in trials, and in our setting with a higher live birth rate per transfer. This cohort alone cannot establish a causal advantage. Women who need HRT-FET because of anovulation or irregular cycles should be told of the higher hypertensive risk and may benefit from closer blood pressure monitoring. Whether low-dose aspirin lowers this risk in HRT-FET pregnancies deserves testing in a trial. Estimates within natural cycle subtypes were imprecise but pointed the same way.

5. Conclusion

In this single-center cohort, HRT-FET was associated with a lower live birth rate per transfer and a two-fold higher risk of hypertensive disorders of pregnancy than natural cycle FET, with a slightly higher birthweight and no significant difference in standardized growth categories or prematurity within the precision of the study. The direction agrees with recent randomised trials and is compatible with the corpus luteum hypothesis. The findings support natural cycle preparation as the first option when ovulation can be monitored.

Acknowledgments

The authors thank the clinical and laboratory staff of American Hospital Kosovo for their patient care and record-keeping.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency.

Ethics approval and informed consent

The study was approved by the Kosovo Doctors Chamber Ethics Committee (reference 153/2025, 12 May 2025) and conducted in accordance with the Declaration of Helsinki. All patients had given written informed consent at the start of treatment for the use of their de-identified data in research; no additional consent was required for this retrospective analysis.

Data availability

The dataset is not publicly available because of institutional policy and patient privacy. It is available from the corresponding author on reasonable request with institutional approval. The analysis code is available on request.

Author contributions

GG designed the study, extracted and analyzed the data, and wrote the first draft. GAY contributed to interpreting the results and to writing and critically revising the manuscript. VM and TIH verified flagged records against source documents, contributed to data collection and interpretation, and revised the manuscript critically. All authors read and approved the final manuscript and agree to be accountable for the work.

Use of artificial intelligence

The authors used Claude (Anthropic) and ChatGPT/Codex (OpenAI) for language editing and typo corrections between June and October 2026. The authors reviewed and verified all content and take full responsibility for the manuscript.

References

1. Roque M, Haahr T, Geber S, Esteves SC, Humaidan P. Fresh versus elective frozen embryo transfer in IVF/ICSI cycles: a systematic review and meta-analysis of reproductive outcomes. Human Reproduction Update 2019; 25 (1): 2-14. https://doi.org/10.1093/humupd/dmy033

2. Chen ZJ, Shi Y, Sun Y, Zhang B, Liang X et al. Fresh versus frozen embryos for infertility in the polycystic ovary syndrome. New England Journal of Medicine 2016; 375 (6): 523-533. https://doi.org/10.1056/NEJMoa1513873

3. Mackens S, Santos-Ribeiro S, van de Vijver A, Racca A, Van Landuyt L et al. Frozen embryo transfer: a review on the optimal endometrial preparation and timing. Human Reproduction 2017; 32 (11): 2234-2242. https://doi.org/10.1093/humrep/dex285

4. Conrad KP, Baker VL. Corpus luteal contribution to maternal pregnancy physiology and outcomes in assisted reproductive technologies. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 2013; 304 (2): R69-R72. https://doi.org/10.1152/ajpregu.00239.2012

5. Conrad KP. Maternal vasodilation in pregnancy: the emerging role of relaxin. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 2011; 301 (2): R267-R275. https://doi.org/10.1152/ajpregu.00156.2011

6. Singh B, Reschke L, Segars J, Baker VL. Frozen-thawed embryo transfer: the potential importance of the corpus luteum in preventing obstetrical complications. Fertility and Sterility 2020; 113 (2): 252-257. https://doi.org/10.1016/j.fertnstert.2019.12.007

7. von Versen-Höynck F, Schaub AM, Chi YY, Chiu KH, Liu J et al. Increased preeclampsia risk and reduced aortic compliance with in vitro fertilization cycles in the absence of a corpus luteum. Hypertension 2019; 73 (3): 640-649. https://doi.org/10.1161/HYPERTENSIONAHA.118.12043

8. Ginström Ernstad E, Wennerholm UB, Khatibi A, Petzold M, Bergh C. Neonatal and maternal outcome after frozen embryo transfer: increased risks in programmed cycles. American Journal of Obstetrics and Gynecology 2019; 221 (2): 126.e1-126.e18. https://doi.org/10.1016/j.ajog.2019.03.010

9. Saito K, Kuwahara A, Ishikawa T, Morisaki N, Miyado M et al. Endometrial preparation methods for frozen-thawed embryo transfer are associated with altered risks of hypertensive disorders of pregnancy, placenta accreta, and gestational diabetes mellitus. Human Reproduction 2019; 34 (8): 1567-1575. https://doi.org/10.1093/humrep/dez079

10. Asserhøj LL, Spangmose AL, Aaris Henningsen AK, Clausen TD, Ziebe S et al. Adverse obstetric and perinatal outcomes in 1,136 singleton pregnancies conceived after programmed frozen embryo transfer (FET) compared with natural cycle FET. Fertility and Sterility 2021; 115 (4): 947-956. https://doi.org/10.1016/j.fertnstert.2020.10.039

11. Busnelli A, Schirripa I, Fedele F, Bulfoni A, Levi-Setti PE. Obstetric and perinatal outcomes following programmed compared to natural frozen-thawed embryo transfer cycles: a systematic review and meta-analysis. Human Reproduction 2022; 37 (7): 1619-1641. https://doi.org/10.1093/humrep/deac073

12. Moreno-Sepulveda J, Espinós JJ, Checa MA. Lower risk of adverse perinatal outcomes in natural versus artificial frozen-thawed embryo transfer cycles: a systematic review and meta-analysis. Reproductive BioMedicine Online 2021; 42 (6): 1131-1145. https://doi.org/10.1016/j.rbmo.2021.03.002

13. Epelboin S, Labrosse J, De Mouzon J, Devaux A, Gervoise-Boyer MJ et al. Higher risk of pre-eclampsia and other vascular disorders with artificial cycle for frozen-thawed embryo transfer compared to ovulatory cycle or to fresh embryo transfer following in vitro fertilization. Frontiers in Endocrinology 2023; 14: 1182148. https://doi.org/10.3389/fendo.2023.1182148

14. Wang B, Zhang J, Zhu Q, Yang X, Wang Y. Effects of different cycle regimens for frozen embryo transfer on perinatal outcomes of singletons. Human Reproduction 2020; 35 (7): 1612-1622. https://doi.org/10.1093/humrep/deaa093

15. Zong L, Liu P, Zhou L, Wei D, Ding L et al. Increased risk of maternal and neonatal complications in hormone replacement therapy cycles in frozen embryo transfer. Reproductive Biology and Endocrinology 2020; 18 (1): 36. https://doi.org/10.1186/s12958-020-00601-3

16. Groenewoud ER, Cohlen BJ, Al-Oraiby A, Brinkhuis EA, Broekmans FJM et al. A randomized controlled, non-inferiority trial of modified natural versus artificial cycle for cryo-thawed embryo transfer. Human Reproduction 2016; 31 (7): 1483-1492. https://doi.org/10.1093/humrep/dew120

17. Ho VNA, Pham TD, Nguyen NT, Wang R, Norman RJ et al. Livebirth rate after one frozen embryo transfer in ovulatory women starting with natural, modified natural, or artificial endometrial preparation in Viet Nam: an open-label randomised controlled trial. The Lancet 2024; 404 (10449): 266-275. https://doi.org/10.1016/S0140-6736(24)00756-6

18. Geysenbergh B, Debrock S, Lie Fong S, Vanhie A, Tomassetti C et al. Natural versus artificial cycle for endometrial preparation in ovulatory women undergoing frozen-thawed embryo transfer: a randomised controlled trial. Human Reproduction 2026; 41 (8). https://doi.org/10.1093/humrep/deag095

19. Wei D, Qin Y, Sun Y, Yan J, Zhao H et al. Natural ovulation versus programmed regimens before frozen embryo transfer in ovulatory women: multicentre, randomised clinical trial. BMJ 2026; 392: e087045. https://doi.org/10.1136/bmj-2025-087045

20. Liu X, Li W, Wen W, Wang T, Wang T et al. Natural cycle versus hormone replacement therapy as endometrial preparation in ovulatory women undergoing frozen-thawed embryo transfer: the COMPETE open-label randomized controlled trial. PLOS Medicine 2025; 22 (6): e1004630. https://doi.org/10.1371/journal.pmed.1004630

21. Zaat TR, Kostova EB, Korsen P, Showell MG, Mol F et al. Obstetric and neonatal outcomes after natural versus artificial cycle frozen embryo transfer and the role of luteal phase support: a systematic review and meta-analysis. Human Reproduction Update 2023; 29 (5): 634-654. https://doi.org/10.1093/humupd/dmad011

22. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010; 21 (3): 383-388. https://doi.org/10.1097/EDE.0b013e3181d61eeb

23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet 2007; 370 (9596): 1453-1457. https://doi.org/10.1016/S0140-6736(07)61602-X

24. American College of Obstetricians and Gynecologists. Gestational hypertension and preeclampsia: ACOG Practice Bulletin, Number 222. Obstetrics and Gynecology 2020; 135 (6): e237-e260. https://doi.org/10.1097/AOG.0000000000003891

25. International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33 (3): 676-682. https://doi.org/10.2337/dc09-1848

26. Villar J, Cheikh Ismail L, Victora CG, Ohuma EO, Bertino E 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. The Lancet 2014; 384 (9946): 857-868. https://doi.org/10.1016/S0140-6736(14)60932-6

27. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. Journal of the American Statistical Association 2018; 113 (521): 390-400. https://doi.org/10.1080/01621459.2016.1260466

28. Thomas LE, Li F, Pencina MJ. Overlap weighting: a propensity score method that mimics attributes of a randomized clinical trial. JAMA 2020; 323 (23): 2417-2418. https://doi.org/10.1001/jama.2020.7819

29. Zou G. A modified Poisson regression approach to prospective studies with binary data. American Journal of Epidemiology 2004; 159 (7): 702-706. https://doi.org/10.1093/aje/kwh090

30. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Annals of Internal Medicine 2017; 167 (4): 268-274. https://doi.org/10.7326/M16-2607

31. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research 2011; 46 (3): 399-424. https://doi.org/10.1080/00273171.2011.568786

32. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. Journal of Statistical Software 2011; 42 (8): 1-28. https://doi.org/10.18637/jss.v042.i08

33. Zeileis A. Object-oriented computation of sandwich estimators. Journal of Statistical Software 2006; 16 (9): 1-16. https://doi.org/10.18637/jss.v016.i09

34. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004; 15 (5): 615-625. https://doi.org/10.1097/01.ede.0000135174.63482.43

35. Levine RJ, Maynard SE, Qian C, Lim KH, England LJ et al. Circulating angiogenic factors and the risk of preeclampsia. New England Journal of Medicine 2004; 350 (7): 672-683. https://doi.org/10.1056/NEJMoa031884

36. Nelissen ECM, van Montfoort APA, Dumoulin JCM, Evers JLH. Epigenetics and the placenta. Human Reproduction Update 2011; 17 (3): 397-417. https://doi.org/10.1093/humupd/dmq052

37. Labarta E, Mariani G, Holtmann N, Celada P, Remohí J et al. Low serum progesterone on the day of embryo transfer is associated with a diminished ongoing pregnancy rate in oocyte donation cycles after artificial endometrial preparation: a prospective study. Human Reproduction 2017; 32 (12): 2437-2442. https://doi.org/10.1093/humrep/dex316

Supplementary tables

Supplementary Table S1. Endometrial preparation protocol by year of transfer, all analyzed cycles.

Year	Cycles	HRT-FET	HRT-FET, %
2012	214	214	100.0
2013	221	221	100.0
2014	238	237	99.6
2015	274	266	97.1
2016	249	39	15.7
2017	221	32	14.5
2018	231	37	16.0
2019	255	28	11.0
2020	186	23	12.4
2021	225	22	9.8
2022	228	24	10.5
2023	221	42	19.0
2024	180	23	12.8

HRT-FET accounted for 100% of cycles in 2012–2013, 99.6% in 2014, and 97.1% in 2015.

Supplementary Table S2. Maternal and neonatal outcomes in the 1:1 propensity score matched cohort (188 pairs, n = 376).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Maternal outcomes					
Hypertensive disorders of pregnancy	27/188 (14.4)	7/188 (3.7)	3.86 (1.68–8.88)	0.002	+10.6 (+4.7 to +16.5)
Pre-eclampsia	19/188 (10.1)	6/188 (3.2)	3.17 (1.26–7.95)	0.014	+6.9 (+1.8 to +12.1)
Gestational diabetes	17/188 (9.0)	19/188 (10.1)	0.89 (0.47–1.69)	0.732	−1.1 (−7.2 to +5.0)
Postpartum hemorrhage	13/188 (6.9)	10/188 (5.3)	1.30 (0.57–2.97)	0.534	+1.6 (−3.4 to +6.6)
Cesarean delivery	84/188 (44.7)	86/188 (45.7)	0.98 (0.78–1.23)	0.839	−1.1 (−11.3 to +9.2)
Neonatal outcomes					
Small for gestational age	7/188 (3.7)	5/187 (2.7)	1.39 (0.44–4.39)	0.572	+1.0 (−2.6 to +4.7)
Large for gestational age	29/188 (15.4)	23/187 (12.3)	1.25 (0.74–2.12)	0.400	+3.1 (−4.1 to +10.4)
Low birthweight (<2,500 g)	5/188 (2.7)	7/188 (3.7)	0.71 (0.23–2.26)	0.567	−1.1 (−4.7 to +2.6)
Macrosomia (>4,000 g)	22/188 (11.7)	16/188 (8.5)	1.37 (0.73–2.58)	0.321	+3.2 (−3.1 to +9.5)
Preterm birth (<37 weeks)	12/188 (6.4)	15/188 (8.0)	0.80 (0.41–1.57)	0.515	−1.6 (−6.4 to +3.2)
NICU admission	18/188 (9.6)	13/188 (6.9)	1.38 (0.69–2.77)	0.357	+2.7 (−3.0 to +8.3)
Birthweight, g	3410 ± 495	3309 ± 492	+101 (+1 to +201)	0.047	—
Gestational age at delivery, weeks	39.4 ± 1.6	39.1 ± 1.8	+0.24 (−0.09 to +0.57)	0.147	—

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Apgar score at 5 min	9.6 ± 0.7	9.6 ± 0.7	+0.02 (−0.11 to +0.15)	0.809	—
Rare outcomes used as negative controls					
Congenital anomaly	4/188 (2.1)	2/188 (1.1)	2.00 (0.36–10.97)	0.425	+1.1 (−1.5 to +3.6)
Neonatal death	0/188 (0.0)	2/188 (1.1)	NE	—	−1.1 (−2.5 to +0.4)

Nearest-neighbor matching without replacement within a caliper of 0.2 SD of the logit propensity score; standard errors clustered by matched pair. Growth outcomes use 375 births (one natural cycle birth outside the INTERGROWTH-21st range). MD, mean difference; NE, not estimable (no events in one group).

Supplementary Table S3. Each natural cycle subtype compared with HRT-FET in separate 1:1 matched analyses.

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Natural vs HRT (123 pairs)					
Hypertensive disorders	16/123 (13.0)	6/123 (4.9)	2.67 (1.04–6.84)	0.041	+8.1 (+0.8 to +15.5)
Pre-eclampsia	13/123 (10.6)	4/123 (3.3)	3.25 (1.05–10.01)	0.040	+7.3 (+0.8 to +13.8)
Small for gestational age	7/123 (5.7)	6/123 (4.9)	1.17 (0.43–3.19)	0.764	+0.8 (−4.5 to +6.1)
Preterm birth	6/123 (4.9)	12/123 (9.8)	0.50 (0.20–1.26)	0.143	−4.9 (−11.2 to +1.5)
Low birthweight	4/123 (3.3)	3/123 (2.4)	1.33 (0.30–5.99)	0.708	+0.8 (−3.4 to +5.1)
Modified natural vs HRT (149 pairs)					
Hypertensive disorders	20/149 (13.4)	8/149 (5.4)	2.50 (1.10–5.69)	0.029	+8.1 (+1.2 to +14.9)
Pre-eclampsia	15/149 (10.1)	8/149 (5.4)	1.87 (0.79–4.44)	0.152	+4.7 (−1.6 to +11.0)
Small for gestational age	7/149 (4.7)	6/149 (4.0)	1.17 (0.39–3.48)	0.782	+0.7 (−4.1 to +5.4)
Preterm birth	10/149 (6.7)	10/149 (6.7)	1.00 (0.43–2.30)	1.000	+0.0 (−5.6 to +5.6)
Low birthweight	5/149 (3.4)	8/149 (5.4)	0.63 (0.20–1.92)	0.411	−2.0 (−6.8 to +2.7)
Letrozole vs HRT (97 pairs)					
Hypertensive disorders	10/97 (10.3)	6/97 (6.2)	1.67 (0.60–4.61)	0.325	+4.1 (−4.0 to +12.2)
Pre-eclampsia	8/97 (8.2)	2/97 (2.1)	4.00 (0.84–18.99)	0.081	+6.2 (−0.1 to +12.5)
Small for gestational age	5/97 (5.2)	5/97 (5.2)	1.00 (0.29–3.48)	1.000	+0.0 (−6.4 to +6.4)
Preterm birth	6/97 (6.2)	8/97 (8.2)	0.75 (0.26–2.17)	0.596	−2.1 (−9.7 to +5.5)
Low birthweight	4/97 (4.1)	4/97 (4.1)	1.00 (0.25–4.03)	1.000	+0.0 (−5.8 to +5.8)
Clomiphene vs HRT (112 pairs)					

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Hypertensive disorders	15/112 (13.4)	8/112 (7.1)	1.87 (0.79–4.44)	0.153	+6.2 (−2.1 to +14.6)
Pre-eclampsia	12/112 (10.7)	5/112 (4.5)	2.40 (0.84–6.84)	0.102	+6.2 (−0.9 to +13.4)
Small for gestational age	6/112 (5.4)	2/112 (1.8)	3.00 (0.60–14.97)	0.180	+3.6 (−1.4 to +8.5)
Preterm birth	8/112 (7.1)	10/112 (8.9)	0.80 (0.33–1.93)	0.619	−1.8 (−8.8 to +5.3)
Low birthweight	4/112 (3.6)	3/112 (2.7)	1.33 (0.30–6.00)	0.708	+0.9 (−3.8 to +5.6)

Each subtype was matched separately to HRT-FET births using the same propensity score specification.

Supplementary Table S4. Spontaneous and modified natural cycles only (mild stimulation cycles excluded) versus HRT-FET, 1:1 matched analysis (176 pairs).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Hypertensive disorders of pregnancy	22/176 (12.5)	10/176 (5.7)	2.20 (1.06–4.55)	0.033	+6.8 (+0.8 to +12.9)
Pre-eclampsia	16/176 (9.1)	9/176 (5.1)	1.78 (0.78–4.03)	0.169	+4.0 (−1.6 to +9.5)
Gestational diabetes	19/176 (10.8)	17/176 (9.7)	1.12 (0.58–2.15)	0.740	+1.1 (−5.6 to +7.8)
Postpartum hemorrhage	12/176 (6.8)	10/176 (5.7)	1.20 (0.52–2.78)	0.671	+1.1 (−4.1 to +6.4)
Cesarean delivery	84/176 (47.7)	88/176 (50.0)	0.95 (0.77–1.18)	0.671	−2.3 (−12.8 to +8.2)
Small for gestational age	7/176 (4.0)	7/176 (4.0)	1.00 (0.35–2.86)	1.000	+0.0 (−4.2 to +4.2)
Large for gestational age	26/176 (14.8)	25/176 (14.2)	1.04 (0.63–1.70)	0.876	+0.6 (−6.6 to +7.7)
Macrosomia (>4,000 g)	19/176 (10.8)	18/176 (10.2)	1.06 (0.58–1.91)	0.858	+0.6 (−5.7 to +6.8)
NICU admission	15/176 (8.5)	11/176 (6.2)	1.36 (0.67–2.80)	0.397	+2.3 (−3.0 to +7.5)
Preterm birth (<37 weeks)	11/176 (6.2)	12/176 (6.8)	0.92 (0.42–2.01)	0.828	−0.6 (−5.7 to +4.6)
Low birthweight (<2,500 g)	6/176 (3.4)	4/176 (2.3)	1.50 (0.42–5.33)	0.531	+1.1 (−2.4 to +4.7)
Congenital anomaly	3/176 (1.7)	3/176 (1.7)	1.00 (0.20–4.98)	1.000	+0.0 (−2.7 to +2.7)
Neonatal death	0/176 (0.0)	3/176 (1.7)	NE	—	−1.7 (−3.6 to +0.2)

Supplementary Table S5. Calendar-time and model-specification sensitivity analyses: restriction to 2016–2024, era-stratified estimates, and alternative propensity model specifications (overlap weighting).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Per cycle, 2016–2024 (n = 1,996)					
Positive β-hCG	144/270 (53.5)	1098/1726 (63.0)	0.85 (0.75–0.96)	0.008	−9.5 (−16.1 to −3.0)
Clinical pregnancy	136/270 (50.6)	1043/1726 (59.8)	0.85 (0.74–0.96)	0.010	−9.2 (−15.8 to −2.6)
Live birth	113/270 (41.9)	901/1726 (51.3)	0.82 (0.70–0.95)	0.010	−9.4 (−16.0 to −2.8)

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Singleton live births, 2016–2024 (n = 956)					
Hypertensive disorders	13/109 (12.1)	63/847 (6.7)	1.81 (1.01–3.24)	0.045	+5.4 (−1.0 to +11.9)
Pre-eclampsia	11/109 (10.1)	45/847 (5.0)	2.02 (1.05–3.87)	0.035	+5.1 (−0.8 to +11.0)
Gestational diabetes	11/109 (10.2)	89/847 (10.5)	0.97 (0.53–1.77)	0.928	−0.3 (−6.4 to +5.8)
Postpartum hemorrhage	7/109 (6.6)	49/847 (5.7)	1.15 (0.52–2.51)	0.734	+0.8 (−4.2 to +5.9)
Cesarean delivery	54/109 (49.3)	406/847 (48.8)	1.01 (0.82–1.25)	0.914	+0.6 (−9.7 to +10.8)
Small for gestational age	6/109 (5.7)	34/846 (3.4)	1.71 (0.72–4.07)	0.226	+2.4 (−2.3 to +7.1)
Large for gestational age	16/109 (14.7)	117/846 (12.7)	1.16 (0.71–1.88)	0.561	+2.0 (−5.1 to +9.0)
Low birthweight	4/109 (3.9)	33/847 (3.1)	1.25 (0.44–3.55)	0.674	+0.8 (−3.2 to +4.7)
Macrosomia	12/109 (11.1)	85/847 (9.6)	1.16 (0.65–2.09)	0.616	+1.5 (−4.8 to +7.9)
Preterm birth	6/109 (5.6)	69/847 (7.6)	0.73 (0.32–1.66)	0.454	−2.0 (−6.8 to +2.7)
NICU admission	10/109 (9.3)	51/847 (6.8)	1.36 (0.71–2.59)	0.357	+2.4 (−3.3 to +8.2)
Hypertensive disorders by era (weights re-estimated within era)					
2012–2017	41/409 (11.9)	8/196 (2.4)	4.92 (1.60–15.07)	0.005	+9.5 (+2.1 to +16.8)
2018–2024	10/80 (12.6)	55/656 (7.4)	1.71 (0.88–3.33)	0.112	+5.3 (−2.5 to +13.1)
Hypertensive disorders, alternative propensity model specifications (all singleton live births)					
PS without endometrial thickness	51/489 (12.3)	63/852 (5.7)	2.18 (1.33–3.57)	0.002	+6.6 (+2.1 to +11.2)
PS with year as 3 categories	51/489 (12.0)	63/852 (6.3)	1.90 (1.08–3.35)	0.026	+5.7 (−0.5 to +11.9)
Year-adjusted weighted outcome model	—	—	2.11 (1.30–3.40)	—	—

Weights were re-estimated within each restricted sample or era. In 2016–2024 the propensity model c-statistic was 0.67. The 2012–2017 estimate rests on 8 natural cycle events and is unstable. These analyses were added after an independent methodological audit and were not part of the original analysis plan.

Supplementary Table S6. Term singleton births (≥ 37 weeks), 1:1 matched exploratory analysis (177 pairs).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Hypertensive disorders	24/177 (13.6)	6/177 (3.4)	4.00 (1.63–9.81)	0.002	+10.2 (+4.3 to +16.1)
Pre-eclampsia	15/177 (8.5)	5/177 (2.8)	3.00 (1.09–8.28)	0.034	+5.6 (+0.7 to +10.6)

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Small for gestational age	5/177 (2.8)	2/177 (1.1)	2.50 (0.48–12.95)	0.275	+1.7 (−1.2 to +4.6)
Large for gestational age	29/177 (16.4)	20/177 (11.3)	1.45 (0.85–2.48)	0.174	+5.1 (−2.2 to +12.3)
Low birthweight	2/177 (1.1)	3/177 (1.7)	0.67 (0.11–4.01)	0.658	−0.6 (−3.1 to +1.9)
Macrosomia	21/177 (11.9)	13/177 (7.3)	1.62 (0.81–3.23)	0.175	+4.5 (−1.9 to +11.0)
NICU admission	14/177 (7.9)	7/177 (4.0)	2.00 (0.81–4.97)	0.135	+4.0 (−1.1 to +9.0)
Birthweight, g	3460 ± 452	3377 ± 444	+83 (−16 to +183)	0.101	—

This analysis conditions on gestational age at delivery and is hypothesis generating.

Supplementary Table S7. Placental complications among cesarean deliveries with documented placental findings (n = 265 of 619), overlap-weighted exploratory analysis.

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Placenta previa	3/93 (4.7)	1/172 (0.2)	25.47 (2.53–256.91)	0.006	+4.5 (−1.3 to +10.4)
Placental abruption	2/93 (1.9)	1/172 (0.2)	7.87 (0.70–88.74)	0.095	+1.7 (−1.1 to +4.5)
Uterine atony	0/93 (0.0)	6/172 (1.1)	NE	—	−1.1 (−2.4 to +0.2)

Event numbers are very small; estimates are imprecise.

Supplementary Table S8. Interaction between protocol and single versus double embryo transfer, overlap-weighted models.

Outcome	Ratio of HRT-FET risk ratios, double vs single embryo transfer	P for interaction
Hypertensive disorders	1.14	0.828
Pre-eclampsia	1.08	0.919
Small for gestational age	0.18	0.080
Preterm birth	1.11	0.866

A ratio above 1 means the HRT-FET risk ratio was larger after double embryo transfer.

Supplementary Table S9. Growth outcomes under the clinical-database flags instead of the INTERGROWTH-21st definition (overlap weighting, all singleton live births).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Small for gestational age (database flag)	28/489 (4.2)	63/852 (9.0)	0.47 (0.25–0.87)	0.016	−4.8 (−8.4 to −1.1)
Large for gestational age (database flag)	62/489 (12.6)	94/852 (10.5)	1.21 (0.79–1.84)	0.379	+2.2 (−2.8 to +7.1)

The database flags were of unrecorded reference and disagreed with the sex-specific INTERGROWTH-21st classification in 136 (SGA) and 314 (LGA) of 1,340 births. They are shown only to document the sensitivity of the growth results to the definition.

Supplementary Table S2b. Matched analysis with inference clustered by patient-linked groups of pairs (154 components from 188 pairs; 27 women contributed to more than one pair).

Outcome	HRT-FET, n/N (%)	Natural cycle, n/N (%)	RR (95% CI)	P	Risk difference, pp (95% CI)
Hypertensive disorders	27/188 (14.4)	7/188 (3.7)	3.86 (1.72–8.66)	0.001	+10.6 (+4.9 to +16.4)
Pre-eclampsia	19/188 (10.1)	6/188 (3.2)	3.17 (1.26–7.95)	0.014	+6.9 (+1.8 to +12.1)
Small for gestational age	7/188 (3.7)	5/187 (2.7)	1.39 (0.44–4.39)	0.572	+1.0 (−2.6 to +4.7)
Large for gestational age	29/188 (15.4)	23/187 (12.3)	1.25 (0.75–2.11)	0.392	+3.1 (−4.0 to +10.3)

Compare with Supplementary Table S2, which clusters by pair.

Appendix A. Changes from version 1

Hypertensive disorders of pregnancy after hormone replacement therapy versus natural cycle frozen embryo transfer: a propensity score-weighted cohort study

Goc G, Aynaoğlu Yıldız G, Maloku V, Isufaj Haliti T. American Hospital Kosovo and University Clinical Centre of Kosovo, Prishtina; Ankara University Faculty of Medicine, Ankara, Türkiye.

Version 1: Research Square, DOI 10.21203/rs.3.rs-8437063/v1.

Why a new version

Version 1 contained numbers that did not agree with each other (cohort counts differed between the Methods, the flow diagram and the tables; several risk ratios did not reproduce from the reported event counts) and two references that were cited for findings they do not contain. The dataset was re-checked, the analysis was rewritten as a single reproducible script, and an independent methodological audit of the revised package was obtained before this version was finalised. Version 2 replaces every number in version 1. The study design, setting, period and exposure definition are unchanged. The changes fall into seven separate groups, listed below so that each can be judged on its own.

1. Changes to the data file

Comparison of the version-1 data file with the current file shows that the two contain the same 3,000 cycles, the same 70 variables and the same patient keys. Cell-level differences are confined to three columns: 45 free-text entries in "other maternal complications" and 80 in "other neonatal complications" were cleared, and 8 perinatal mortality flags were changed from 1 to 0. Seven of these eight flags were in cycles without a positive pregnancy test and one in a cycle that ended in miscarriage; none was in a live birth. No covariate had missing values in either file. Version 1 stated that 147 cycles were excluded for missing covariates; we have not been able to reproduce that statement from either file and now regard it as a reporting error in version 1.

A second verification step followed: a screen of the file identified 243 records with internally inconsistent values (infants flagged both small and large for gestational age, preterm flags with recorded gestational age of 37 weeks or more, nulliparous women with a previous cesarean delivery, HRT cycles with endometrial thickness below 7 mm, and maternal age decreasing between cycles). Two clinicians checked these against source documents. 268 values in 201 patients were corrected (maternal age 97, preterm flag 61, endometrial thickness 43, nulliparity 28, transfer year 24, SGA 10, LGA 5) and infant sex was added; each field-level correction is recorded in a change log, and the verifying clinician and source document for each entry are held by the authors. No logical inconsistency remains; the analysis code halts if one appears. Plurality is known only from the fetal heart count at the first scan, so vanishing twins cannot be identified; this is now stated as a limitation.

2. Change to the analytical cohort (an exclusion, not a data correction)

Version 2 excludes 57 cycles that were converted from a natural cycle to HRT after failed follicular development, as the Methods of version 1 had stated but the analysis had not done. This exclusion alone accounts for the change from 3,000 to 2,943 cycles, from 1,457 to 1,439 live births, and from 1,358 to 1,341 singleton live births (the 57 cycles contained 18 live births, 17 of them singletons).

3. Change to the primary analysis

Overlap weighting on all 1,341 singleton live births is now the primary analysis; in version 1 it was a robustness analysis and 1:1 matching was primary. The reason is that 1:1 matching retained only 188 of 489 HRT-FET births and could not balance transfer year below a standardized difference of 0.10, whereas overlap weighting uses all births and balances every covariate in the model; it also targets the clinical-equipoise population that version 1 had already described as its estimand. Matching is kept as the confirmatory analysis and both are reported side by side.

Other analytical changes: the propensity model is described correctly as 20 variables (26 model terms), not "29 covariates"; standardized mean differences for binary variables are now computed with the pooled standard deviation rather than as raw proportion differences; miscarriage is reported per clinical pregnancy rather than per positive test; deaths within the first 28 days after live birth are called neonatal deaths; and absolute risk differences carry confidence intervals. The following analyses were added after the audit and are labelled as such in the Methods: restriction to 2016–2024, era-stratified estimates with weights re-estimated within each era, a year-adjusted weighted model, E-values, and two alternative propensity model specifications.

4. Change to the perinatal death result (denominator, not only data)

Version 1 reported perinatal death as a negative control with RR 0.12 (P = 0.004). That analysis used all 3,000 cycles as the denominator and included the eight death flags described in section 1. Version 2 analyses neonatal death (within 28 days of birth) among singleton live births, where the counts are identical in the old and new files (2 HRT-FET vs 5 natural cycle), giving RR 0.18 (95% CI 0.03–1.10). The disappearance of statistical significance therefore reflects both the corrected flags and the change of denominator; it is not evidence of no effect, and the estimate remains very imprecise.

5. Gestational age units and growth definition (second review)

A second, computational review found that gestational age was recorded as completed weeks and days (37.6 = 37 weeks 6 days) and had been read as decimal weeks. All analyses now convert it to days. Preterm birth is unaffected (the recorded flag and the recalculated flag agree in every birth); the continuous gestational-age estimate changed slightly.

The same review recalculated small and large for gestational age from birthweight, infant sex and exact gestational day using the INTERGROWTH-21st newborn standards. The database flags disagreed with the standard in 136 (SGA) and 314 (LGA) of 1,340 births, and the lower SGA rate after HRT-FET that version 2 had reported with the database flags (RR 0.47) disappears under the standard (RR 1.21, 95% CI 0.57–2.59; LGA RR 1.25, 0.86–1.81). The manuscript now uses the INTERGROWTH-21st definition for all growth outcomes and shows the database-flag results only as a sensitivity table. Birthweight (+89 g) is unchanged. A matched analysis clustered by patient-linked groups of pairs, added because 27 women appear in more than one pair, gave the same estimates as pair clustering.

6. Changes to the text

- The headline is now a two-fold, not 3.5-fold, increase in hypertensive disorders.
- The introduction and discussion were rewritten around the randomised trials published since 2016 (ANTARCTICA, Ho 2024, COMPETE 2025, PnROVE 2026, Geysenbergh 2026) and the 2023 meta-analysis of obstetric outcomes. The paper is now framed as an independent real-world confirmation, not as filling an evidence gap.

- Reference 19 of version 1 was cited as the ANTARCTICA trial but was a meta-analysis on FSH triggering; reference 6 was cited as the French registry but was a meta-analysis. A draft of version 2 cited ANTARCTICA-2, which has no HRT arm, for a natural-versus-artificial comparison; this has been corrected to ANTARCTICA-1. The list now has 37 references, each checked against its Crossref record.
- Claims of "largest effect size reported", "strong support for the corpus luteum hypothesis", "no adverse effects on neonatal outcomes" and "should be preferred" were removed or narrowed. Neonatal outcomes are described as showing no significant difference within the precision of the study.
- The calendar-time structure of the cohort (HRT-FET in 97–100% of cycles before 2016) is reported as the main design limitation.
- The "term birth paradox" section was shortened and is reported as exploratory.
- Software versions are reported as actually used; the figures are stated to be drawn from the exported tables.
- The abstract was shortened to under 300 words and the keywords reduced to six.

7. Change to authorship

Gülşah Aynaoğlu Yıldız (Ankara University) has been added as second author for her contributions to the interpretation of the results and to writing and critical revision of the manuscript. The author order is now Goksu Goc, Gülşah Aynaoğlu Yıldız, Velsa Maloku, Tefta Isufaj Haliti. All four authors have consented to the change in writing.

Summary of main estimates, version 1 versus version 2

Result	Version 1	Version 2
Live birth per transfer, RR	0.59 (0.46–0.76)	0.82 (0.74–0.92)
Miscarriage, RR	1.43 (0.94–2.16), per positive test	1.41 (1.03–1.93), per clinical pregnancy
Hypertensive disorders, primary RR	3.50 (1.41–8.70), 164 matched pairs	2.15 (1.31–3.54), weighted, n = 1,341
Hypertensive disorders, matched RR	—	3.86 (1.68–8.88), 188 pairs
Pre-eclampsia, primary RR	3.20 (1.17–8.76)	2.08 (1.14–3.80)
Absolute risk difference, HDP	+12.2 points	+6.5 points (2.0–11.1)
Hypertensive disorders, 2016–2024 only	not reported	1.81 (1.01–3.24)
Birthweight, mean difference	+34 g (P = 0.49)	+89 g (12–166), P = 0.024
Small for gestational age, RR	0.69 (0.31–1.56), database flags	1.21 (0.57–2.59), INTERGROWTH-21st; 0.47 (0.25–0.87) with database flags (sensitivity)
Death as negative control	0.12 (P = 0.004), 3,000 cycles	0.18 (0.03–1.10), 1,341 singleton live births
E-value for HDP	not reported	3.7 (CI bound 1.9)

Gestational diabetes, postpartum hemorrhage, cesarean delivery, preterm birth, low birthweight and NICU admission show no significant difference in either version. Under the standardized growth definition, no significant difference was detected for small or large for gestational age.

Authors' note

We apologise for the errors in version 1. All four authors have reviewed and approved this note and the revised manuscript. The analysis code that produced every number in version 2, and the record-level change log, are available from the corresponding author.

Figures

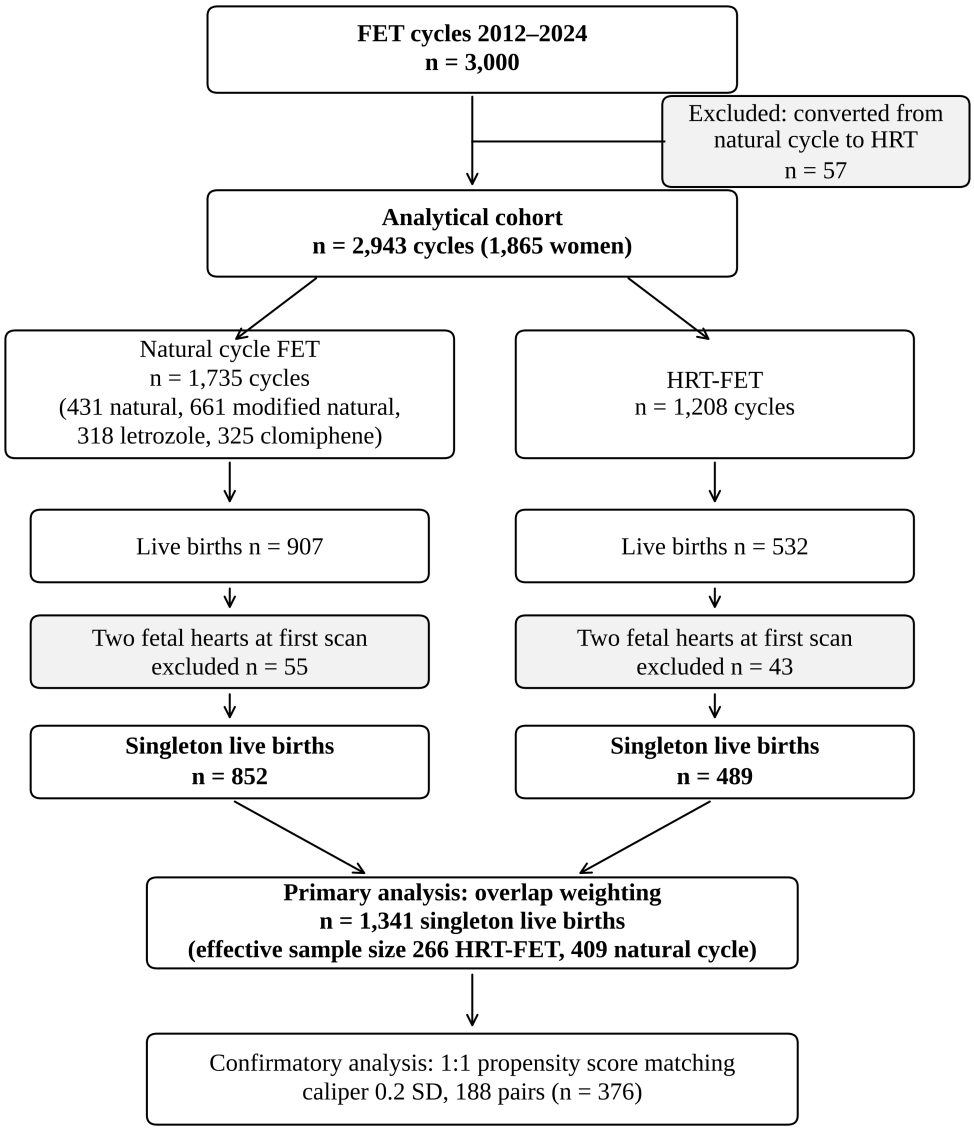


Figure 1

Figure 1. Flow of cycles through the study. HRT, hormone replacement therapy; FET, frozen embryo transfer.

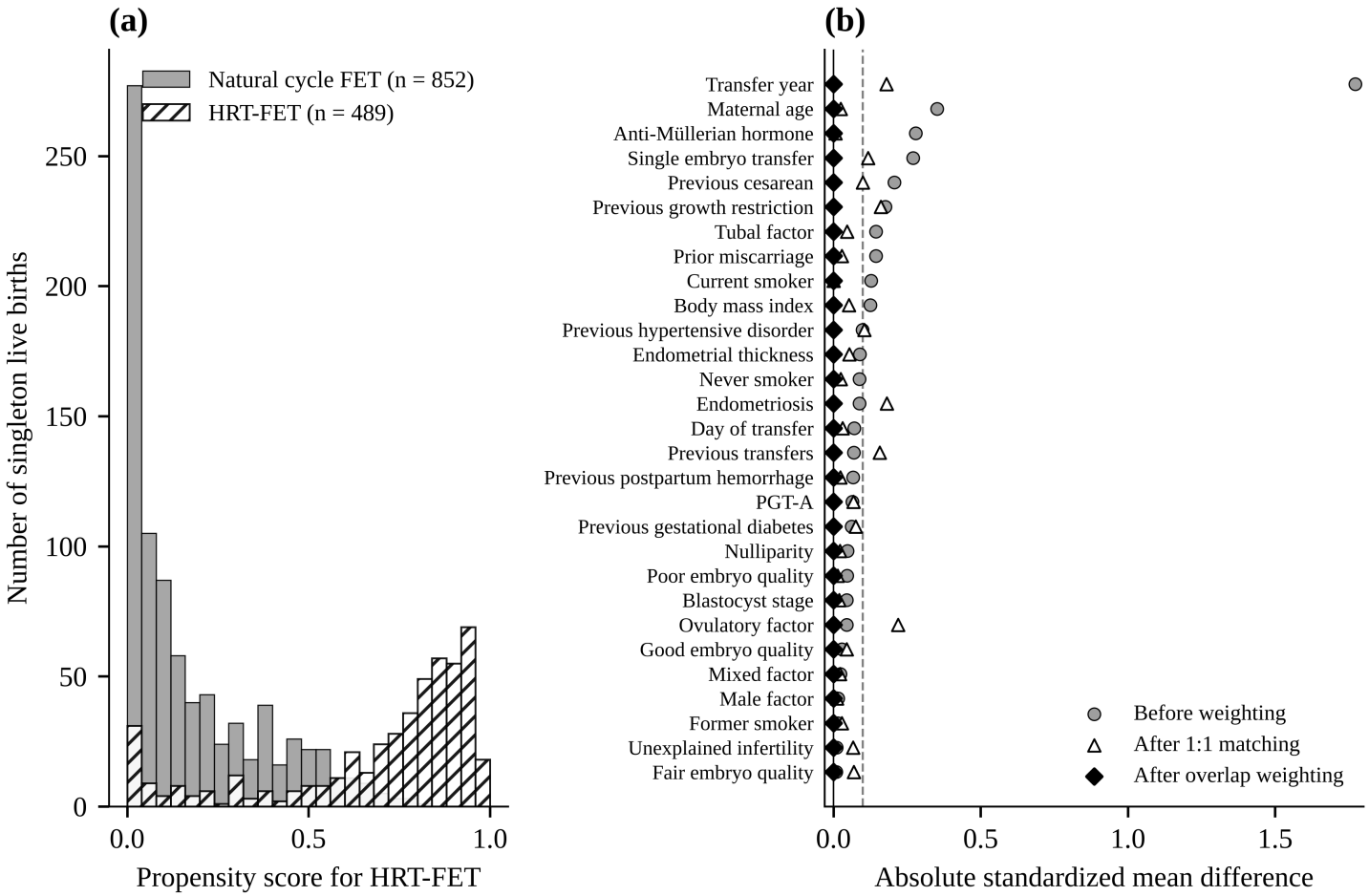


Figure 2

Figure 2. Propensity score diagnostics in singleton live births. (a) Propensity score distributions by protocol before weighting. (b) Absolute standardized mean differences before weighting (grey circles), after 1:1 matching (open triangles), and after overlap weighting (black diamonds), all computed from the analytic data and weights; the dashed line marks 0.10. Standardized mean differences use the pooled standard deviation for binary and continuous variables.

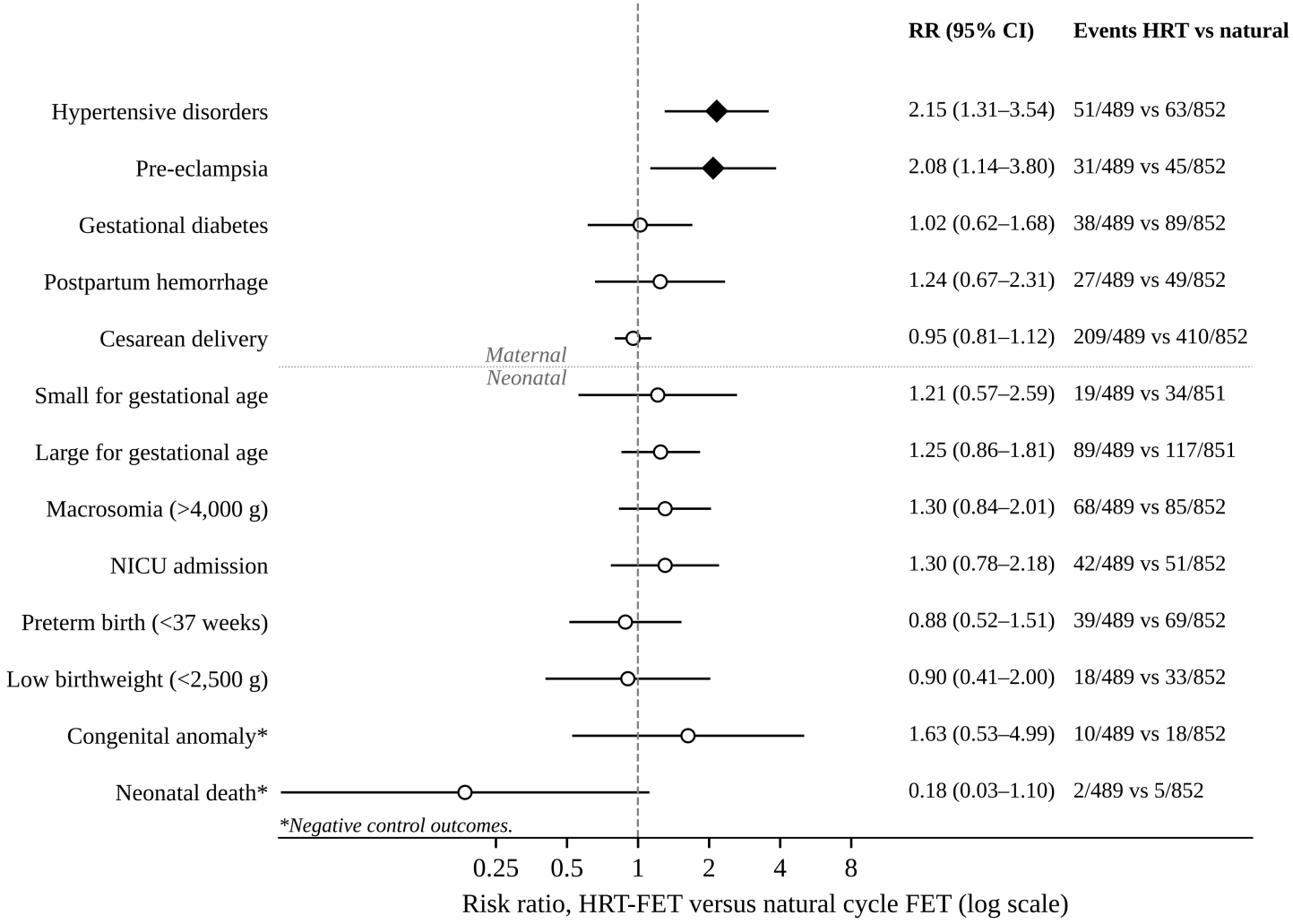


Figure 3

Figure 3. Risk ratios (95% confidence intervals) for maternal and neonatal outcomes, HRT-FET versus natural cycle FET, from the overlap-weighted analysis of 1,341 singleton live births (1,340 for small and large for gestational age, INTERGROWTH-21st definition). Values above 1 indicate higher risk with HRT-FET.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx