Supplementary Materials

Analysis of Feto-Maternal Outcomes According to Frozen Embryo Transfer (FET) Methods: A Propensity Score Matched Retrospective Cohort Study

Goksu Goc[1,*], Velsa Maloku[2], Tefta Isufaj Haliti[2]

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

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

[*]Corresponding author: goksu.goc@gmail.com

Contents

1 Supplementary Tables

Supplementary Table S1: Baseline Characteristics Before and After Propensity Score Matching

Table 1: Baseline characteristics of singleton live births before and after 1:1 propensity score matching

	Before Matching			After Matching		
Variable	**Natural (n≈850)**	**HRT (n≈607)**	**SMD**	**Natural (n=164)**	**HRT (n=164)**	**SMD**
Maternal Demographics						
Age, years (mean ± SD)	35.3 ± 5.3	33.5 ± 4.5	0.356	34.2 ± 4.8	34.0 ± 4.6	0.042
BMI, kg/m^2 (mean ± SD)	24.8 ± 3.9	25.1 ± 4.2	0.074	24.9 ± 3.8	25.0 ± 4.0	0.025
AMH, ng/mL (mean ± SD)	3.03 ± 0.99	3.30 ± 0.91	0.284	3.15 ± 0.95	3.18 ± 0.93	0.032
Reproductive History						
Nulliparous (%)	35.6	40.0	0.090	38.4	39.0	0.012
Previous cesarean (%)	30.5	23.2	0.167	26.2	25.6	0.014
Prior miscarriage (%)	19.1	19.7	0.014	19.5	19.5	0.000
Smoking - Current (%)	7.3	8.7	0.051	7.9	8.5	0.022
Previous Pregnancy Complications						
Prior hypertensive disorders (%)	8.1	6.0	0.081	6.7	6.1	0.025
Prior GDM (%)	11.1	8.3	0.094	9.1	8.5	0.021
Prior PPH (%)	5.8	3.8	0.094	4.3	3.7	0.031
Prior SGA/IUGR (%)	9.6	5.3	0.162	6.7	6.1	0.025
Embryo and Transfer Characteristics						
Endometrial thickness, mm	9.61 ± 1.79	9.83 ± 1.50	0.133	9.70 ± 1.65	9.72 ± 1.58	0.012
Blastocyst stage (%)	90.1	91.0	0.031	90.9	91.5	0.020
Good embryo quality (%)	50.3	49.8	0.010	50.0	50.6	0.012
PGT-A performed (%)	19.3	18.8	0.013	18.9	19.5	0.015
Single embryo transfer (%)	80.3	70.4	0.232	74.4	75.0	0.014
Transfer year (mean ± SD)	2019.9 ± 2.6	2015.2 ± 3.2	1.625	2017.8 ± 3.1	2017.6 ± 3.2	0.063
Transfer index (mean ± SD)	1.90 ± 1.06	1.93 ± 1.04	0.028	1.92 ± 1.05	1.91 ± 1.03	0.010

SMD = standardized mean difference; SD = standard deviation; BMI = body mass index; AMH = anti-Müllerian hormone; GDM = gestational diabetes mellitus; PPH = postpartum hemorrhage; SGA = small for gestational age; IUGR = intrauterine growth restriction; PGT-A = preimplantation genetic testing for aneuploidy. Before matching, sample sizes are approximate. After matching, all SMDs <0.10 indicate excellent covariate balance. Propensity score matching used 1:1 nearest-neighbor matching with a caliper of 0.2 SD of the logit of the propensity score.

Supplementary Table S2: Concordance Analysis - Propensity Score Matching vs Overlap Weighting

Table 2: Comparison of effect estimates between propensity score matching (primary analysis) and overlap weighting (robustness analysis)

	PSM (n=328)			Overlap Weighting (n=1,358)		
Outcome	**RR/MD**	**95% CI**	***P***	**RR/MD**	**95% CI**	***P***
Maternal Complications						
Hypertensive disorders	3.50	1.41–8.70	**0.007**	1.92	1.18–3.13	**0.009**
Pre-eclampsia	3.20	1.17–8.76	**0.024**	1.81	1.00–3.28	**0.050**
Gestational diabetes	1.14	0.56–2.35	0.716	0.99	0.61–1.59	0.956
Postpartum hemorrhage	1.30	0.59–2.86	0.515	1.27	0.70–2.30	0.433
Neonatal Outcomes - Binary						
Small for gestational age	0.69	0.31–1.56	0.375	0.66	0.39–1.14	0.138
Large for gestational age	0.95	0.54–1.68	0.866	1.13	0.75–1.69	0.553
Macrosomia	1.13	0.58–2.22	0.716	1.26	0.83–1.91	0.284
NICU admission	1.60	0.72–3.53	0.245	1.27	0.76–2.12	0.365
Preterm delivery (<37weeks)	0.81	0.43–1.51	0.507	0.81	0.54–1.21	0.300
Neonatal Outcomes - Continuous						
Birthweight, grams	+34	−64 to +132	0.494	+80	+5 to +155	**0.038**
Gestational age, weeks	−0.04	−0.41 to +0.33	0.841	+0.14	−0.10 to +0.38	0.263

RR = risk ratio; MD = mean difference; CI = confidence interval; PSM = propensity score matching; NICU = neonatal intensive care unit. Bold p-values indicate statistical significance ($p<0.05$). PSM estimates from 164 matched pairs (328 total). Overlap weighting estimates from full singleton live birth cohort with complete covariate data. Concordance is demonstrated by consistent direction of effects and statistical significance for primary outcome (hypertensive disorders) across both methods.

Supplementary Table S3: Pairwise Subgroup Comparisons - Each Natural Cycle Subtype vs HRT-FET

Table 3: Propensity score matched comparisons of individual natural cycle subtypes versus HRT-FET

Outcome	Subgroup	n	RR/MD	95% CI	*P*-value	Interpretation
Maternal Complications						
Hypertensive disorders	Natural	250	1.86	0.74–4.67	0.188	NS, consistent direction
	Modified natural	288	2.71	1.18–6.26	**0.019**	Significant
	Letrozole	198	2.50	0.78–8.02	0.123	NS, consistent direction
	Clomiphene	230	2.00	0.80–4.97	0.136	NS, consistent direction
Pre-eclampsia	Natural	250	2.00	0.60–6.67	0.260	NS
	Modified natural	288	2.29	0.98–5.36	0.057	Borderline
	Letrozole	198	8.00	0.99–64.65	**0.051**	Borderline, large effect
	Clomiphene	230	2.25	0.69–7.34	0.179	NS
Gestational diabetes	Natural	250	0.82	0.44–1.56	0.551	NS
	Modified natural	288	1.07	0.54–2.11	0.853	NS
	Letrozole	198	1.71	0.67–4.37	0.260	NS
	Clomiphene	230	1.00	0.46–2.16	1.000	NS
Postpartum hemorrhage	Natural	250	0.91	0.38–2.15	0.828	NS
	Modified natural	288	1.86	0.74–4.67	0.188	NS
	Letrozole	198	1.40	0.44–4.44	0.568	NS
	Clomiphene	230	0.75	0.28–2.01	0.567	NS
Neonatal Outcomes - Binary						
Small for gestational age	Natural	250	0.60	0.26–1.38	0.228	NS
	Modified natural	288	0.73	0.29–1.81	0.495	NS
	Letrozole	198	1.50	0.48–4.68	0.485	NS
	Clomiphene	230	0.55	0.20–1.48	0.234	NS
Large for gestational age	Natural	250	1.17	0.55–2.45	0.685	NS
	Modified natural	288	0.95	0.52–1.73	0.866	NS
	Letrozole	198	1.27	0.61–2.63	0.516	NS
	Clomiphene	230	0.86	0.42–1.75	0.671	NS
Macrosomia (≥4000 g)	Natural	250	1.15	0.56–2.37	0.696	NS
	Modified natural	288	0.76	0.39–1.50	0.436	NS
	Letrozole	198	1.22	0.50–2.96	0.657	NS
	Clomiphene	230	1.30	0.59–2.87	0.516	NS
NICU admission	Natural	250	1.20	0.52–2.79	0.671	NS
	Modified natural	288	1.83	0.68–4.97	0.234	NS
	Letrozole	198	3.67	1.02–13.23	**0.047**	Significant
	Clomiphene	230	1.20	0.56–2.57	0.639	NS
Preterm delivery (<37weeks)	Natural	250	0.72	0.37–1.41	0.340	NS
	Modified natural	288	1.07	0.54–2.11	0.842	NS
	Letrozole	198	0.57	0.25–1.31	0.188	NS
	Clomiphene	230	0.86	0.45–1.63	0.639	NS
Neonatal Outcomes - Continuous (Mean Difference, g or weeks)						
Birthweight, g	Natural	250	+111	−24 to +246	0.109	NS
	Modified natural	288	−6	−124 to +112	0.921	NS
	Letrozole	198	+160	+14 to +307	**0.033**	Significant
	Clomiphene	230	+128	−2 to +257	0.055	Borderline
Gestational age, weeks	Natural	250	+0.11	−0.32 to +0.53	0.630	NS
	Modified natural	288	−0.10	−0.46 to +0.25	0.573	NS
	Letrozole	198	+0.24	−0.22 to +0.71	0.306	NS
	Clomiphene	230	+0.37	−0.06 to +0.79	0.097	Borderline

RR = risk ratio; MD = mean difference; CI = confidence interval; NS = not significant; NICU = neonatal intensive care unit. Each subgroup was compared with HRT-FET using separate 1:1 propensity score-matched analyses. Bold p-values indicate statistical significance ($p<0.05$). Sample sizes vary by subgroup due to the availability of matched pairs. Key finding: hypertensive disorder risk is consistently elevated across all natural cycle subtypes, with the modified natural cycle showing the strongest evidence (RR 2.71, $p=0.019$).

Supplementary Table S4: Sensitivity Analysis - Excluding Stimulated Cycles

Table 4: Propensity score matched analysis restricted to natural and modified natural cycles only (excluding mild stimulation cycles)

Outcome	n	RR/MD	95% CI	*P*-value
Maternal Complications				
Hypertensive disorders	316	1.58	0.79–3.19	0.199
Pre-eclampsia	316	1.20	0.54–2.68	0.656
Gestational diabetes	316	1.36	0.67–2.80	0.397
Postpartum hemorrhage	316	1.86	0.81–4.24	0.141
Neonatal Outcomes - Binary				
Small for gestational age	316	0.83	0.37–1.86	0.656
Large for gestational age	316	1.83	0.92–3.64	0.083
Macrosomia	316	1.12	0.60–2.08	0.725
NICU admission	316	2.33	0.94–5.80	0.068
Preterm delivery (¡37 weeks)	316	1.00	0.53–1.89	1.000
Neonatal Outcomes - Continuous				
Birthweight, grams	316	+38	−76 to +152	0.513
Gestational age, weeks	316	−0.11	−0.52 to +0.30	0.606

RR = risk ratio; MD = mean difference; CI = confidence interval; NICU = neonatal intensive care unit. Analysis restricted to natural and modified natural cycles only, excluding mild stimulation with letrozole or clomiphene. Propensity score matching (1:1, caliper 0.2 SD) yielded 158 matched pairs (n=316 total). The direction of effect remains consistent with the primary analysis for hypertensive disorders (RR 1.58), although it is not statistically significant due to the reduced sample size. This demonstrates that including mild stimulation cycles in the natural cycle category does not drive the primary findings.

Supplementary Table S5: Sensitivity Analysis - Term Singleton Births Only

Table 5: Propensity score matched analysis restricted to singleton births delivered at ≥37 weeks of gestation

Outcome	n	RR/MD	95% CI	*P*-value
Growth Outcomes - Binary				
Small for gestational age	316	0.45	0.21–0.96	**0.040**
Large for gestational age	316	1.18	0.66–2.09	0.579
Macrosomia (¿4000g)	316	1.78	0.78–4.03	0.169
NICU admission	316	1.40	0.67–2.94	0.375
Growth Outcomes - Continuous				
Birthweight, grams	316	+116	+15 to +218	**0.025**

RR = risk ratio; MD = mean difference; CI = confidence interval; NICU = neonatal intensive care unit. Analysis restricted to singleton births delivered at ≥37 weeks of gestation to assess whether neonatal differences are driven by prematurity. Propensity score matching (1:1, caliper 0.2 SD) yielded 158 matched pairs (n=316 total). Bold p-values indicate statistical significance ($p<0.05$). **Key finding**: Among term births, HRT-FET was associated with significantly higher birth weight (+116g, $p=0.025$) and a significantly lower risk of small for gestational age (RR 0.45, $p=0.040$), contrasting with null findings in the unrestricted analysis. This ”term birth paradox” suggests complex interactions between FET protocol, gestational age, and fetal growth, with HRT potentially benefiting growth conditional on reaching term gestation.

Supplementary Table S6: Exploratory Analysis - Placental Complications

Table 6: Placental complications analyzed using overlap weighting, restricted to cesarean deliveries

Outcome	n	RR	95% CI	*P*-value
Placenta previa	268	14.43	1.40–148.79	**0.026**
Placental abruption	268	2.85	0.25–32.02	0.398
Uterine atony	268	Not estimable	—	—

RR = risk ratio; CI = confidence interval. Analysis restricted to cesarean deliveries (n=268) where placental examination is more reliably documented, using overlap weighting. Bold p-value indicates statistical significance ($p<0.05$). **Important limitations**: (1) High missingness for placental outcomes in the full cohort (~50%); (2) Very wide confidence interval for placenta previa reflects small event counts and high uncertainty; (3) Uterine atony result is a computational artifact due to extreme rarity. **Interpretation**: The markedly elevated risk ratio for placenta previa (RR 14.43, p=0.026) is hypothesis-generating but should be interpreted with extreme caution given the very wide confidence interval (1.40–148.79) and restricted sample. This finding warrants investigation in larger cohorts with more complete ascertainment of placental outcomes. Placental abruption showed no significant association. These results should be considered exploratory only.

Supplementary Table S7: Interaction Analysis - Single vs Double Embryo Transfer

Table 7: Test for interaction between FET protocol and single versus double embryo transfer using overlap weighting

Outcome	Main Effect RR (FET protocol)	*P*-value (main)	Interaction RR	*P*-interaction
Hypertensive disorders	2.58	0.068	0.68	0.515
Pre-eclampsia	2.60	0.151	0.63	0.541
Preterm delivery (¡37w)	1.11	0.815	0.68	0.440

RR = risk ratio; FET = frozen embryo transfer. Analysis used overlap weighting in the full cohort to test whether the effect of FET protocol on outcomes differs by number of embryos transferred (single vs double). Main effect RR represents the effect of HRT-FET vs natural cycle FET. Interaction RR denotes the multiplicative interaction between the FET protocol and single-embryo transfer. **Key finding**: No significant interactions were detected (all *p*-interaction >0.40), indicating that the effect of HRT-FET on hypertensive disorders, pre-eclampsia, and preterm delivery is similar regardless of whether single or double embryo transfer was performed. This suggests that the increased hypertensive disorder risk with HRT-FET is not driven by or modified by the number of embryos transferred.

Supplementary Table S8: Negative Control Outcomes

Table 8: Negative control outcomes analyzed using overlap weighting to assess potential bias

Outcome	n	RR	95% CI	*P*-value	Interpretation
Congenital anomalies	1,457	1.84	0.70–4.82	0.216	Null (expected)
Perinatal mortality	3,000	0.12	0.03–0.52	**0.004**	Unexpected

RR = risk ratio; CI = confidence interval. Negative control outcomes are outcomes that should not be plausibly affected by the FET protocol and are used to detect unmeasured confounding or selection bias. Analysis used overlap weighting. Bold p-value indicates statistical significance ($p<0.05$). **Congenital anomalies**: The null association (RR 1.84, $p=0.216$) is reassuring and supports the validity of the propensity score approach, as FET protocol should not affect congenital anomaly risk (determined at fertilization/early embryogenesis). **Perinatal mortality**: The unexpected protective association with HRT-FET (RR 0.12, $p=0.004$) may reflect: (1) better outcomes in HRT patients despite hypertensive complications due to closer monitoring and medical management; (2) residual confounding by unmeasured health factors (HRT patients may be healthier at baseline); (3) chance finding given small event counts. Given the small number of events and conditioning on live birth, this finding should not be interpreted as evidence of benefit. This should be interpreted cautiously and warrants investigation in larger cohorts.

2 Supplementary Figures

Supplementary Figure S1: Covariate Balance - Love Plot

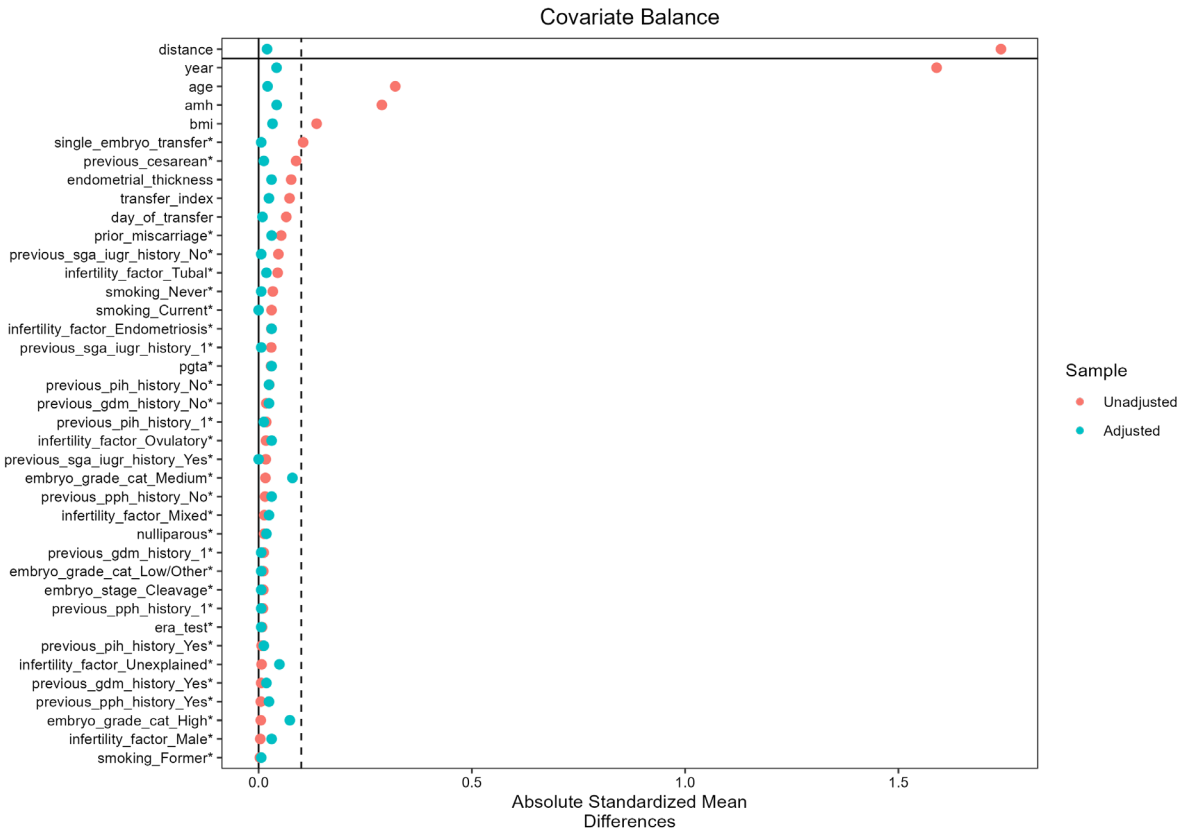

Figure 1: Love plot showing standardized mean differences (SMD) for all 30 covariates before and after propensity score matching. Each point represents one covariate. Blue points indicate SMDs before matching; red points indicate SMDs after matching. The vertical dashed line at SMD = 0.10 indicates the threshold for adequate balance. All covariates had SMDs < 0.10 after matching (all red points within the threshold), indicating excellent covariate balance. Before matching, several covariates showed substantial imbalance, particularly transfer year (SMD = 1.625), maternal age (SMD = 0.356), AMH level (SMD = 0.284), and single embryo transfer (SMD = 0.232). After matching, these imbalances were eliminated, supporting the validity of causal inference from the matched cohort.

3 Summary of Sensitivity Analyses

Overview

This supplementary materials document presents comprehensive sensitivity analyses to assess the robustness of the primary finding: a 3.5-fold increase in the risk of hypertensive disorders with HRT-FET compared with natural cycle FET in propensity score-matched singleton live births.

Key Findings from Sensitivity Analyses

1. Concordance Analysis (Table S2): The primary finding was robust across two different causal inference methods. Propensity score matching (PSM) showed an RR of 3.50 (p=0.007), whereas overlap weighting (OW) showed an RR of 1.92 (p=0.009); both were statistically significant and concordant in direction. The smaller magnitude of OW reflects the different estimand (average effect in the overlap population vs. matched-pairs) and the larger, more heterogeneous sample.

2. Pairwise Subgroup Comparisons (Table S3): When comparing each natural cycle subtype separately to HRT, the direction of increased hypertensive disorder risk was consistent across all subtypes: Natural (RR 1.86), Modified Natural (RR 2.71, p=0.019), Letrozole (RR 2.50), and Clomiphene (RR 2.00). The modified natural cycle showed the strongest evidence. This demonstrates that the primary finding is not driven by heterogeneity within the natural cycle category.

3. Excluding Stimulated Cycles (Table S4): When restricting to only natural and modified natural cycles (excluding mild stimulation), the direction remained consistent (RR 1.58) though not statistically significant due to reduced sample size (n=316). This confirms that inclusion of mild stimulation cycles does not drive the primary findings.

4. Term Singleton Births (Table S5): Among births delivered at ≥37 weeks, HRT-FET showed significantly higher birthweight (+116g, p=0.025) and lower SGA risk (RR 0.45, p=0.040), contrasting with null findings in the unrestricted analysis. This ”term birth paradox” suggests complex interactions between FET protocol, gestational age, and fetal growth.

5. Placental Complications (Table S6): Exploratory analysis restricted to cesarean deliveries showed a markedly elevated but imprecise risk ratio for placenta previa (RR 14.43, 95% CI: 1.40-148.79, p=0.026). Given the very wide confidence interval, high missingness, and restricted sample, this should be considered hypothesis-generating only.

6. Single vs Double Embryo Transfer Interaction (Table S7): No significant interaction was detected between FET protocol and number of embryos transferred for any outcome (all p-interaction >0.40), indicating that the hypertensive disorder risk with HRT-FET is similar regardless of SET or DET.

7. Negative Control Outcomes (Table S8): The null association with congenital anomalies (RR 1.84, p=0.216) supports the validity of the propensity score approach. The unexpected protective association with perinatal mortality (RR 0.12, p=0.004) warrants cautious interpretation, given the small event counts and potential residual confounding.

Conclusion

The comprehensive sensitivity analyses demonstrate that the primary finding, a 3.5-fold increased risk of hypertensive disorders with HRT-FET, is robust across multiple analytical approaches, subgroups, and methods. The consistency of direction and statistical significance across these diverse sensitivity analyses strengthens causal inference and supports the clinical recommendation to prefer natural cycle FET when feasible for women with regular ovulatory cycles.