

**CHRONIC INFLAMMATION AND OXIDATIVE STRESS IN
ENDOMETRIOSIS-ASSOCIATED INFERTILITY**

Xolboyeva Mashhura Yoqutjon qizi

Email – yoqutjonovnam@gmail.com

Termez University of Economics and Service

Faculty of Medicine

ANNOTATION

Background: Endometriosis-associated infertility is a heterogeneous syndrome in which distorted pelvic anatomy interacts with immune dysregulation, iron-dependent oxidative injury, altered follicular and peritoneal environments, and endometrial receptivity. Objective: To critically integrate mechanistic and clinical evidence and examine whether real-world analytics can improve decisions without overstating the validity of candidate biomarkers. Methods: This integrative narrative review prioritizes the 2022 ESHRE endometriosis guideline, recent systematic reviews, randomized trials, and regulatory guidance on real-world evidence and responsible artificial intelligence. It explains confounding, outcome selection, time-zero alignment, propensity scores, inverse probability weighting, target trial emulation, and survival models as methods suitable for future research rather than analyses conducted here. Results: Biological evidence supports reciprocal interactions between macrophage activation, proinflammatory cytokines, recurrent bleeding, free iron, reactive oxygen species, lipid peroxidation, and mitochondrial dysfunction. These pathways plausibly impair oocyte competence, sperm function, embryo development and implantation, but causal effects on live birth are not established by biomarker associations. A recent antioxidant meta-analysis pooled 21 randomized trials and 1,626 participants; the reported clinical pregnancy effect was not significant (risk ratio 1.12, 95% confidence interval 0.79–1.61), despite some favorable symptom and biochemical outcomes. Imaging, stage, age, ovarian reserve, tubal factors, partner fertility and preferences remain essential determinants of care. Real-world datasets can address longer-term outcomes and equity only when exposure, timing, missingness, selection and measurement bias are addressed. Conclusions: Redox and immune mechanisms offer plausible therapeutic targets and useful research hypotheses; they do not currently justify routine antioxidant prescribing to improve conception. Prospective multi-center registries linked to reproductive outcomes, ethically governed computational research, and adequately powered fertility trials are priority next steps. This article distinguishes

mechanistic plausibility from clinical effectiveness and proposals from empirical findings.

Keywords: endometriosis; infertility; oxidative stress; reactive oxygen species; iron overload; macrophages; cytokines; implantation; follicular microenvironment; real-world evidence; target trial emulation; artificial intelligence; reproductive outcomes

ABSTRACT

Background: Endometriosis-associated infertility is a heterogeneous syndrome in which distorted pelvic anatomy interacts with immune dysregulation, iron-dependent oxidative injury, altered follicular and peritoneal environments, and endometrial receptivity. **Objective:** To critically integrate mechanistic and clinical evidence and examine whether real-world analytics can improve decisions without overstating the validity of candidate biomarkers. **Methods:** This integrative narrative review prioritizes the 2022 ESHRE endometriosis guideline, recent systematic reviews, randomized trials, and regulatory guidance on real-world evidence and responsible artificial intelligence. It explains confounding, outcome selection, time-zero alignment, propensity scores, inverse probability weighting, target trial emulation, and survival models as methods suitable for future research rather than analyses conducted here. **Results:** Biological evidence supports reciprocal interactions between macrophage activation, proinflammatory cytokines, recurrent bleeding, free iron, reactive oxygen species, lipid peroxidation, and mitochondrial dysfunction. These pathways plausibly impair oocyte competence, sperm function, embryo development and implantation, but causal effects on live birth are not established by biomarker associations. A recent antioxidant meta-analysis pooled 21 randomized trials and 1,626 participants; the reported clinical pregnancy effect was not significant (risk ratio 1.12, 95% confidence interval 0.79–1.61), despite some favorable symptom and biochemical outcomes. Imaging, stage, age, ovarian reserve, tubal factors, partner fertility and preferences remain essential determinants of care. Real-world datasets can address longer-term outcomes and equity only when exposure, timing, missingness, selection and measurement bias are addressed. **Conclusions:** Redox and immune mechanisms offer plausible therapeutic targets and useful research hypotheses; they do not currently justify routine antioxidant prescribing to improve conception. Prospective multi-center registries linked to reproductive outcomes, ethically governed computational research, and adequately powered fertility trials are priority next steps. This article distinguishes mechanistic plausibility from clinical effectiveness and proposals from empirical findings.

Keywords: endometriosis; infertility; oxidative stress; reactive oxygen species; iron overload; macrophages; cytokines; implantation; follicular microenvironment; real-world evidence; target trial emulation; artificial intelligence; reproductive outcomes

1. INTRODUCTION

Clinical problem and explanatory limits

Endometriosis is a chronic inflammatory condition with diverse anatomic manifestations and symptom trajectories. Its relationship with infertility is neither universal nor mechanistically singular. Superficial lesions, ovarian endometriomas, deep disease, adhesions and coexisting adenomyosis can represent different causal routes to a common clinical presentation. Some people conceive despite extensive disease, while others experience infertility with comparatively limited visible disease. Such discordance argues against relying on staging alone to infer reproductive prognosis. The relevant causal question is how a particular inflammatory and redox profile changes fecundability or live birth after accounting for age, partner factors, ovarian reserve and treatment pathways. [1,2]

Inflammation and reactive oxygen species

The peritoneal cavity is normally an immunologically active reproductive environment rather than a sterile biochemical compartment. Macrophages, mesothelial cells, lymphocytes, endocrine signals and circulating cells contribute to tissue maintenance. In endometriosis, cyclic bleeding, ectopic tissue persistence and recruitment of innate immune cells may produce sustained cytokine signaling. Reactive oxygen species (ROS) are physiologic signaling molecules at controlled concentrations; pathologic oxidative stress occurs when ROS generation, clearance and repair become imbalanced. Therefore oxidative stress is not synonymous with a high concentration of any single biomarker, and its influence must be interpreted according to tissue, sampling time and functional outcome. [1,3,4]

Real-world data and real-world evidence

Real-world data (RWD) are information routinely generated during healthcare delivery or daily life, such as electronic health records, laboratory feeds, claims, imaging reports and patient-recorded symptoms. Real-world evidence (RWE) is clinically interpretable evidence produced by appropriately analyzing these data. A large RWD dataset is not

automatically reliable evidence: diagnosis coding, encounter patterns, treatment selection and follow-up can distort estimands. For endometriosis-associated infertility, RWE could compare time to pregnancy after clinically reasonable pathways, track adverse effects, and assess differential access to assisted reproduction, but only after patient pathways and causal assumptions are explicitly specified. [6,7]

RCTs, efficacy and effectiveness

Randomized controlled trials address efficacy under protocol-specified assignment and eligibility criteria; pragmatic randomized trials and high-quality observational studies may address effectiveness in more heterogeneous settings. Fertility studies require particular attention to the distinction between biochemical pregnancy, clinical pregnancy, ongoing pregnancy and live birth. An intervention might suppress an inflammatory marker, relieve pelvic pain or increase oocyte yield without increasing live birth. Symptom response and infertility response are therefore separate therapeutic targets. Observational estimates of effectiveness are vulnerable to confounding by indication because patients with more severe disease may receive different interventions and surveillance intensity. [1,5]

Regulatory and ethical environment

The FDA and EMA recognize that carefully characterized real-world data may complement trials in medicines regulation; the EMA DARWIN EU network is an operational example of distributed observational analyses. Neither agency regards observational datasets as a substitute for appropriate causal design. WHO guidance emphasizes transparency, human oversight, equity, protection of privacy and accountability when artificial intelligence is adopted in healthcare. For endometriosis, these standards are particularly salient because fertility information, cycle tracking and reproductive decisions are intimate and potentially stigmatizing. [6–10]

Critical scope and clinical purpose

The present review is organized around an explicit evidence hierarchy. Molecular and animal experiments establish plausibility but cannot directly quantify human live-birth benefits. Diagnostic association studies test discrimination but not whether acting on the marker improves outcomes. Cohorts quantify prognosis or associations but require stronger assumptions for causal effects. Randomized treatment studies provide direct intervention evidence, although their endpoints, size and transportability still require

appraisal. This hierarchy also determines how AI tools and mechanistic predictors should be assessed: model accuracy is insufficient unless predictions lead to better, fairer and safe clinical decisions. [1,3,5]

2. MATERIALS AND METHODS

Design and transparency. This is a structured, critical integrative narrative review rather than a PRISMA-compliant systematic review or meta-analysis. No original participant-level data were obtained, no pooled effect estimates were recalculated, and no new machine-learning model was trained. Accordingly, the results are a thematic synthesis of published findings and explicitly identified research proposals, not a report of new statistical analyses.

Evidence identification. Clinical and mechanistic literature was located through accessible indexed records and publisher pages, with attention to PubMed-indexed research, Cochrane-style evidence synthesis, the ESHRE clinical guideline, WHO, FDA, EMA and relevant methodological scholarship. PubMed, Embase, Scopus, Web of Science and Cochrane Library are recommended databases for a future reproducible systematic search; access to or completed exhaustive searches of every database is not claimed. Journal platforms such as Nature Medicine, The Lancet, JAMA and NEJM were treated as possible publication venues rather than databases whose comprehensive holdings were individually screened.

Inclusion logic. Priority was given to 2020–2026 clinical studies, systematic reviews, major guidelines and mechanistic studies that can contextualize fertility. Earlier seminal papers were retained where necessary for a core mechanism or methodology. Studies focused exclusively on pain were not used to infer reproductive benefit. Mechanistic findings from cell culture and animal models were treated as hypothesis-generating. Papers without independently verifiable bibliographic detail were not represented as confirmed clinical findings.

Synthesis framework. Evidence was organized by compartment—peritoneal cavity, follicular fluid, ovary, endometrium, immune system and embryo—and by the outcome each study could actually support. The methodological evaluation considered selection bias, treatment-selection mechanisms, timing of exposure, outcome definition, attrition, missing data, assay heterogeneity, multiplicity and precision. For observational designs, target-population generalizability was kept separate from internal causal validity.

Propensity score matching

Estimate the conditional probability of treatment from pre-treatment covariates, match patients within adequate overlap, diagnose standardized mean differences and report exclusions. Matching cannot balance unmeasured surgical expertise, disease severity or patient preferences.

Inverse probability weighting

Use inverse treatment probabilities to construct a pseudo-population balanced on measured baseline determinants. Inspect extreme weights, positivity, weight truncation and effective sample size; consider time-varying confounding when treatment changes over fertility follow-up.

Cox proportional hazards regression

Model hazard of a clearly defined event such as first clinical pregnancy, with diagnosis or treatment anchored to a consistent time zero. Check proportional hazards, competing risks and differential observation; a hazard ratio does not directly equal a pregnancy probability.

Kaplan–Meier and survival analysis

Describe cumulative event-free probability when censoring is noninformative. If treatment stops, loss to follow-up or assisted reproduction occurs differentially, standard Kaplan–Meier may misrepresent untreated conception. Competing-risk estimators and restricted mean time-to-event can be useful.

Target trial emulation

Specify eligibility, competing treatment strategies, assignment, time zero, follow-up, outcome, causal contrast and analysis before processing RWD. Avoid immortal-time bias when comparing immediate IVF, surgery-first and expectant management strategies.

Causal inference

Use directed acyclic graphs to distinguish confounders, mediators and colliders. Do not adjust for post-treatment markers when estimating total treatment effects. Compare assumptions through sensitivity analyses for unmeasured confounding and measurement error.

Machine learning and deep learning

Potential prediction tasks include IVF ovarian response or endometrioma recurrence using imaging, longitudinal records and omics. Use patient-level train-validation-test separation, calibration, external validation, decision curves and subgroup audits; cross-validation alone is insufficient.

Natural language processing and large language models

Extract surgical stage, lesion location, fertility intention and adverse effects from de-identified clinical notes with human adjudication. LLM-generated summaries require provenance, accuracy review and checks for unsupported assertions.

Reinforcement learning

Sequential decision strategies for treatment changes could be studied in simulation or carefully designed offline datasets. Exploration on patients is not ethically acceptable without robust safety constraints and approval; confounding and limited policy overlap challenge offline evaluation.

3. RESULTS: THEMATIC EVIDENCE SYNTHESIS

The following results summarize existing mechanistic and clinical evidence. No frequency or numerical estimate below should be read as deriving from a new cohort assembled for this manuscript. The mechanistic links are graded descriptively as established physiologic pathways, associated human findings, or experimental hypotheses rather than combined into an invented numerical score.

3.1 Pelvic inflammatory microenvironment

Repeated activation of innate immune pathways in the peritoneal environment involves macrophage recruitment, cytokine release, prostaglandin signaling and changes in immune-cell phenotype. Macrophage number or polarization alone is not a sufficiently specific infertility test: menstrual phase, coexisting pelvic disease and sampling at surgery introduce variability. IL-1 β , IL-6, TNF- α and chemokine networks may influence gamete transport, ovarian activity and receptivity, but pathway activation does not automatically prove clinically meaningful fertility impairment.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical

pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.2 Macrophage phenotypes and failed resolution

The binary M1/M2 description is a simplification of diverse cellular states. Resident and recruited macrophages may simultaneously participate in clearance of menstrual debris, angiogenesis, lesion survival and tissue remodeling. In endometriosis, defective clearance, reciprocal lesion–macrophage signaling and persistent iron-rich debris could prevent normal inflammatory resolution. Single-cell profiling offers mechanistic detail but studies vary in stage, treatment exposure and sample-processing methods; definitive cell-state signatures with clinical fertility utility remain unvalidated.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and

imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.3 Ectopic bleeding, heme and free iron

Hemoglobin released during repeated bleeding introduces heme and catalytic iron into lesion-adjacent compartments. Iron participates in Fenton chemistry and can amplify oxidative membrane and DNA injury. Iron is essential for normal cellular metabolism, so excessive local exposure rather than iron itself is the relevant biological problem. A 2023 systematic review reported excess iron in endometriotic tissues and linked iron-mediated oxidative stress to relevant disease biology; observational tissue measurements alone cannot prove that iron removal increases conception. [3]

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.4 Ferroptosis and lipid peroxidation

Ferroptosis is an iron-associated, phospholipid-peroxidation-dependent form of regulated cell death. GPX4-related antioxidant defenses, glutathione availability and membrane lipid composition shape cell susceptibility. Endometriotic lesions can display redox adaptations that support survival under stress while nearby healthy reproductive tissues may experience injury. This compartment-specific asymmetry cautions against simplistic claims that increasing antioxidant capacity everywhere will restore fertility; interventions might affect lesions and physiologic ovulatory signaling differently. [4]

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.5 Oxidative stress markers: validity problems

Malondialdehyde, 8-hydroxy-2'-deoxyguanosine, total antioxidant capacity, lipid hydroperoxides and oxidative protein products are used as proxies for redox disturbances. These markers differ in analyte specificity and preanalytic sensitivity; storage temperature, cycle phase, hemolysis and assay kit strongly affect comparability. A change in one marker does not itself validate an intervention. Fertility-specific biomarker qualification requires reproducible measurement, independent prediction beyond standard clinical variables and evidence that acting on the test improves outcomes.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.6 Mitochondria and energetic competence

Mitochondria are central to oocyte maturation, spindle organization, intracellular calcium regulation and early embryonic development. Oxidative injury can disrupt mitochondrial membranes and compromise energy availability, yet mitochondrial ROS are also involved in signaling. Follicular-fluid oxidative measurements are imperfect proxies for individual oocyte competence, and selection into IVF populations constrains generalization to spontaneous conception. Prospective studies should define whether tissue biomarkers improve clinically useful predictions of mature oocyte yield, embryo quality or live birth.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic

signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.7 Follicular fluid and ovarian microenvironment

The follicular compartment is exposed to inflammatory and endocrine inputs distinct from those in peripheral blood. Endometriomas, previous surgery, age and ovarian stimulation protocols can change follicular-fluid constituents. A higher inflammatory marker may reflect underlying ovarian pathology, active stimulation or tissue trauma rather than a direct cause of failed fertilization. Mechanistic fertility research must compare lesions and treatment groups within appropriate clinical strata and avoid conditioning on variables affected by disease and intervention.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured

covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.8 Oocyte maturation and meiotic integrity

Reactive oxygen species can theoretically influence chromosomal segregation, meiotic spindle architecture and cumulus–oocyte communication. However, embryo morphology is a limited surrogate for chromosomal competence and healthy live birth. Age-related aneuploidy, stimulation intensity and gamete factors frequently exert large independent effects. Experimental observations require cautious translation, especially when a study reports cellular endpoints without controlled comparisons of pregnancy or live birth.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.9 Sperm exposure in the inflammatory peritoneal environment

Inflammatory peritoneal fluid can influence sperm motility, capacitation and membrane integrity in ex vivo experimental systems. Such effects are biologically plausible contributors to infertility where gametes encounter altered peritoneal conditions. Their relative importance in natural conception is difficult to estimate because sperm transit, cervical factors, tubal anatomy and partner semen characteristics

co-determine fertility. Prognostic models that omit partner testing risk falsely attributing infertility entirely to endometriosis.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.10 Tubal motility, adhesions and anatomy

Pelvic adhesions and distorted tubo-ovarian relationships are mechanical pathways distinct from molecular inflammation, although inflammation can promote fibrosis. Severe adhesive disease may impair oocyte capture even with normal ovarian reserve. An antioxidant may change redox chemistry but will not necessarily restore an obstructed tube. Analyses that combine patients with minimal disease, deep disease and severe adhesions can mask biologically distinct treatment effects and produce misleading average estimates.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic

signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.11 Endometrial receptivity and progesterone resistance

Eutopic endometrial changes have been implicated in altered implantation signaling, progesterone responsiveness and immune tolerance. Candidate mediators include local prostaglandins, inflammatory transcription factors and altered stromal differentiation. Yet reduced expression of a receptivity marker is not equivalent to implantation failure, and tests must be interpreted in the context of embryo competence. A personalized intervention guided by an endometrial signature should be supported by prospective clinical utility studies rather than association alone.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured

covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.12 Adenomyosis as a frequent competing explanation

Adenomyosis may coexist with endometriosis and is associated with abnormal uterine contractility and reproductive morbidity. It can confound observed links between endometriosis and implantation because diagnostic imaging intensity varies across clinics. Studies that do not report adenomyosis assessment may attribute uterine effects to peritoneal disease. Ideally, transvaginal ultrasound or MRI phenotyping should be standardized and recorded before treatment decisions.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.13 Immune-cell crosstalk at implantation

Natural killer cells, macrophages, T cells and decidual stromal cells coordinate a controlled inflammatory-to-tolerogenic transition around implantation. Immune phenotypes measured in peripheral blood do not necessarily reflect the endometrium. Empiric immunomodulation for a laboratory-defined “immune abnormality” is not justified without validated thresholds, outcome evidence and risk-benefit assessment.

This is especially important in fertility settings where low-risk appearance can obscure meaningful drug exposure during conception.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.14 Pelvic pain, sexual function and conception exposure

Infertility is not exclusively an oocyte–embryo problem. Dyspareunia and chronic pelvic pain can reduce intercourse frequency or timing, while distress, sleep disruption and access constraints affect care pathways. Studies focusing only on biological markers risk excluding behavioral and treatment-mediated factors. A useful causal framework distinguishes physiologic fecundability from the opportunity to conceive, and treats symptom relief as a clinically meaningful endpoint even when it does not prove improved biological fertility.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic

signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.15 RCT evidence for antioxidant supplementation

A 2025 systematic review and meta-analysis included 21 randomized trials totaling 1,626 participants. It identified favorable signals for certain symptoms and malondialdehyde, but did not observe a statistically significant clinical pregnancy benefit: RR 1.12, 95% CI 0.79–1.61, $p = 0.52$. Heterogeneous compounds, doses, comparators and endpoint definitions limit the inference. A null pooled result is not proof that every redox-targeted intervention is ineffective; it means routine fertility benefit has not been demonstrated. [5]

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured

covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

Published trial synthesis	Reported finding			Interpretation	
RCTs/participants	21	trials	/ 1,626	Pooled meta-analysis, heterogeneous compounds	
Clinical pregnancy	RR 1.12; 95% CI 0.79–1.61; p=0.52			Not statistically significant; do not claim fertility benefit	
Oxidative stress	Some	favorable findings	MDA	Biochemical	surrogate, not live birth

3.16 Why a biochemical effect may not produce live birth

The transition from reducing lipid peroxidation to live birth involves multiple causal stages: ovulation, fertilization, tubal transport, implantation, placentation and fetal development. Interventions can improve one intermediate without affecting the total pathway, particularly when anatomical disease, age or male factors remain unresolved. Trials powered for pain or serum biomarkers often cannot rule out clinically important benefit or harm for live birth. Future studies should prespecify live birth and report pregnancy loss, treatment discontinuation and adverse events.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and

pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.17 Clinical management and fertility priorities

The 2022 ESHRE guideline warns against ovarian suppression solely to improve fertility in infertile women with endometriosis. Management should incorporate age, pain, ovarian reserve, endometrioma-related risks, partner evaluation, tubal factors and access to assisted reproductive technology. Surgery may be appropriate for selected indications but can also compromise ovarian reserve, particularly after repeated endometrioma procedures. No mechanistic biomarker currently replaces individualized guideline-based assessment. [1,2]

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.18 Real-world sources: electronic health records

Electronic records can link diagnosis dates, symptoms, surgical findings, imaging, infertility duration and treatment decisions. Their limitations include undocumented

fertility intention, delayed diagnosis and absent out-of-network pregnancies. Failure to capture a live birth may indicate a transfer of care rather than an adverse fertility outcome. Data quality procedures should compare structured codes with narrative and procedure records, establish reliable patient linkage, and audit missing pregnancy and neonatal outcomes.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.19 Real-world sources: prescriptions and pharmacy

Electronic prescribing and dispensing data describe what clinicians ordered and what pharmacies supplied, not necessarily what patients took. Hormonal suppression, analgesic regimens, progestins and fertility medications have different indications and risk periods. Medication exposure should include dose, route, adherence proxies and preconception timing. Confounding by indication is severe when comparing medically managed pain patients with individuals actively attempting conception.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical

population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.20 Real-world sources: laboratories and registries

Laboratory information systems contribute hormone panels, inflammatory markers, oocyte retrieval and fertilization results when linked to ART platforms. Disease and birth registries can provide longer follow-up, although linkage governance and coding quality determine reliability. Biobank specimen collection must record cycle phase, temperature, processing delay and assay batch. Without calibration, combining oxidative-marker values from separate laboratories can generate spurious center effects.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is

unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.21 Real-world sources: pharmacovigilance and claims

Adverse-event reporting systems are designed for safety signal generation, not estimation of comparative incidence without valid denominators. Claims data capture utilization and billing but may incompletely represent clinical staging or reproductive intention. A pharmacovigilance report concerning a supplement or hormonal product cannot be directly interpreted as the frequency of an adverse reaction among users. Linked claims, medical records and registry data are better suited to estimating comparative outcomes after a carefully specified design.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.22 Wearables and patient-facing applications

Cycle-tracking applications and wearables can provide longitudinal pain scores, sleep, activity and self-reported menstrual symptoms between visits. Consumer app participants are selected populations; algorithms can misclassify ovulation and conception timing, and app engagement may correlate with symptoms and

socioeconomic status. Such data should supplement—not replace—validated reproductive endpoints and clinical examination. Consent must specify secondary research use and prevent secondary exploitation of intimate fertility information.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.23 AI: supervised machine learning

Prediction algorithms may combine age, ovarian reserve, ultrasound phenotypes, surgical history and conventional laboratory results to forecast ART response or time to conception. Flexible learners can discover nonlinear interactions but often overfit narrow single-center datasets. Evaluation should report discrimination and calibration, missingness by site, clinical net benefit and transportability. Any improvement over transparent baseline clinical models must be shown with independent data and prospective decision-impact assessment.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for

baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.24 AI: deep imaging and multimodal models

Deep learning could quantify lesion patterns from ultrasound or MRI and combine them with longitudinal reproductive outcomes. Image acquisition, machine settings, operator experience and referral spectrum can dominate performance. A model trained to distinguish surgically confirmed endometriosis from controls may perform poorly when deployed among unselected infertility patients. Multimodal modeling raises additional data-leakage and missing-modality issues that require frozen evaluation protocols.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured

covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.25 AI: NLP, LLMs and reinforcement learning

NLP can recover lesion descriptors and prior treatment histories from unstructured reports, with physician-adjudicated validation. Large language models may assist literature triage or draft summaries but must not invent clinical facts or populate unverified research results. Reinforcement learning is a research avenue for sequential care decisions, not an authorization for autonomous prescribing. Safe offline policy evaluation and independent clinical supervision are prerequisites for any prospective study.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.26 Comparative drug-effectiveness hypotheses

In endometriosis, an effectiveness model should distinguish symptom control from fertility optimization. For example, analgesics may improve quality of life but not directly increase live births, while hormonal suppression can be appropriate for pain yet temporarily prevent conception. Drug-response models developed for diabetes, cardiovascular disease, oncology, infection or neurology offer methodological

analogies—treatment heterogeneity, adverse events, longitudinal covariates—but results from those diseases cannot be transferred as endometriosis-specific effect sizes.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.27 Safety signals and pharmacovigilance

Adverse drug reactions relevant to endometriosis include effects of hormonal treatment, analgesics, procedure-associated drugs and unregulated supplements, each requiring intervention-specific assessment. Drug-safety signal detection tools search for disproportionate reporting, emerging combinations or unusual temporal patterns. They generate hypotheses to investigate, not causal proof. Pregnancy outcomes merit specific surveillance because time-dependent exposure and reporting delays can produce biased safety profiles.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic

signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.28 Precision medicine and biomarker panels

Candidates for personalized care include inflammatory proteins, oxidative products, circulating miRNAs, imaging signatures, ovarian reserve markers and genomic or pharmacogenetic predictors of drug response. Precision medicine is meaningful only if a stratification rule improves decisions compared with existing care. Clinical validation should evaluate assay reproducibility, value beyond age and disease phenotype, decision thresholds, harms of misclassification and whether a biomarker-guided strategy improves a patient-centered endpoint.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured

covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.29 Comparative implications for comorbid diseases

Diabetes, cardiovascular disease, obesity, malignancy, infection and neurologic disorders illustrate why background health status and medication exposure matter in causal research, but their prediction frameworks should not become substitutes for disease-specific biological reasoning. For example, metabolic dysfunction may modify inflammation and conception independently. A model that treats all comorbidities as interchangeable predictors misses distinct causal roles such as treatment contraindication, fertility confounder or competing risk.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.30 Uzbekistan: service integration and prospective research

For Uzbekistan, an ethically governed multi-center reproductive-health registry could link outpatient gynecology, imaging, surgery, ART, pharmacy and obstetric records. The Ministry of Health and relevant digital-health bodies would need to define data stewardship, clinical terminology, opt-out or consent where applicable, lawful linkage and minimum reporting standards. Electronic prescriptions may support drug-exposure

records but should not be equated with adherence. Proposed measures include standardized fertility endpoints, ovarian reserve recording, validated diagnosis codes, data-access review, clinician training and secure linkage with birth outcomes.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.31 External validation and equitable implementation

Transportability requires testing prediction and causal effect estimates in urban and rural services, across language groups and facilities with different diagnostic access. Racial or geographic proxy variables can create inequity if they encode healthcare access instead of disease mechanisms. Missing outcomes may be greatest among financially constrained patients, biasing model calibration. External validation should report subgroup performance and decision consequences rather than only an overall area under the receiver-operating curve.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for

baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

3.32 Evidence synthesis across biological scales

The central synthesis is a multiscale model: recurrent bleeding and iron loading can promote oxidative injury; inflammatory cells amplify local cytokine signaling; ovarian and endometrial environments may become less favorable; and anatomic, endocrine and behavioral pathways interact. Confidence is higher for the presence of inflammatory and oxidative abnormalities than for claims that changing a particular marker increases live birth. Translational priorities should therefore move from biomarker discovery to prespecified interventional hypotheses and clinically meaningful trial outcomes.

Evidence interpretation. This domain must be examined in relation to the exact reproductive estimand: spontaneous conception, oocyte yield, implantation, clinical pregnancy or live birth. The direction and magnitude of associations can differ with disease phenotype, menstrual-cycle phase, surgical exposure and treatment history. Robust appraisal therefore compares the source population with the intended clinical population, specifies when variables were measured, and distinguishes adjustment for baseline confounding from adjustment for downstream mediators. A mechanistic signal should remain a hypothesis until independent clinical data establish predictive and interventional value.

Study-design implications. A feasible prospective program would prespecify eligibility and follow-up, measure age, infertility duration, ovarian reserve, partner factors and imaging phenotype at baseline, then record subsequent treatment changes and pregnancy outcomes. Laboratories should use harmonized specimen handling, blinded assessment and documented quality controls. When an observational comparison is

unavoidable, investigators should define a treatment decision point, evaluate measured covariate overlap, quantify loss to follow-up and conduct sensitivity analyses rather than presenting a single adjusted coefficient as conclusive evidence.

4. DISCUSSION

Mechanistic convergence versus causal certainty

The convergence of human specimen studies, experimental models and redox biology supports the proposition that inflammation and oxidative stress participate in the pathophysiology of endometriosis. Nevertheless, the clinically crucial question is whether modulating one pathway improves time to pregnancy and live birth without unacceptable harms. Lesion persistence, tissue-specific antioxidant adaptation, ovarian reserve and anatomy create competing pathways; consequently, identifying ROS excess is neither sufficient for a fertility diagnosis nor proof of therapeutic reversibility.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Why clinical results disagree

Published studies sample distinct populations and tissues, use different assays and report different outcomes. Patients who undergo surgery or IVF are selected by severity, infertility history and clinical access. A positive association between oxidative markers and poor embryo quality could be confounded by age or stimulation protocol; a negative association could reflect protective treatment given to higher-risk patients. Without careful timing and prespecified estimands, statistical adjustment can aggravate bias by conditioning on mediators or colliders.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Strengths and limitations of artificial intelligence

AI could identify multimodal patterns beyond handcrafted rules, support imaging interpretation and recover structured outcomes from narrative records. But data leakage, model drift, subgroup underrepresentation, poor calibration, shortcut learning and opaque decision rules threaten safety. Explainable AI offers tools to interrogate feature contributions, not a guarantee of biological causation or fairness. Clinical decision support systems should document intended users, action thresholds, oversight, fail-safe behavior and evidence that use improves decisions.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Missing data and measurement error

Missing fertility outcomes are likely informative: individuals achieving pregnancy may stop infertility visits, while those with poor access may disappear from records. Multiple imputation assumes an explicit mechanism and should be complemented by missing-not-at-random sensitivity analyses where warranted. Endometriosis diagnosis misclassification differs by facility because access to imaging and laparoscopy varies. Biomarker error may be heteroscedastic across assays and sampling windows, undermining naive pooling of data.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Ethics and regulatory science

WHO guidance requires protection of autonomy, wellbeing, transparency, accountability and inclusion. FDA RWE guidance and EMA regulatory initiatives encourage reliable observational evidence within rigorous design and data standards.

DARWIN EU illustrates the value of common data models and distributed analytic governance, not the automatic validity of any specific observational endometriosis finding. Algorithmic updates, reproducibility and independent monitoring must be planned as part of clinical implementation.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Implications for Uzbekistan

A useful first stage is to harmonize documentation and build a minimal fertility dataset before purchasing complex AI technology. Standardized patient identifiers, reproductive outcomes, diagnostic phenotype, prescription exposure and consent controls are prerequisites for valid research. Multi-center oversight should involve clinicians, methodologists, information-security experts and patients. Any proposed registry or electronic prescribing linkage must be evaluated against current Uzbek law and actual infrastructure; it should not be described as an already operational nationwide evidence network without verification.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Priorities for prospective research

Researchers should favor pre-registered multicenter cohorts, nested biobanks, assay standardization and pragmatic trials that include live birth as an endpoint. A redox-targeted treatment trial requires justification of dose, timing, biological target, reproductive safety and interaction with established management. Observational target-trial emulation can complement but not replace trials when important residual confounding persists. Economic evaluation should incorporate medication costs, repeated procedures, ART use and opportunity cost to patients.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

Limitations of the present review

This is an integrative review, not a formal exhaustive search or a quantitative synthesis. It does not claim screening counts, pooled meta-analytic effect sizes beyond those attributed to published sources, individual-level regressions or validated AI performance. Regulatory literature is used primarily as a methodological context and is not automatically disease-specific. The breadth requested—including analogies with multiple other diseases—reduces the space available for detailed appraisal of each individual experimental endometriosis study.

A rigorous interpretation separates feasibility from evidence of benefit. Decisions based on findings in one setting require comparison of baseline risk, patient eligibility, measurement methods and competing treatments across settings. Clinically persuasive results should persist after sensitivity analyses and ideally be replicated externally. For fertility, the most defensible intervention endpoint is live birth supplemented by patient-reported outcomes and maternal safety rather than a single molecular surrogate.

5. CONCLUSION

Endometriosis-associated infertility emerges from interacting inflammatory, iron-dependent redox, ovarian, endometrial, anatomic and behavioral mechanisms. These domains are biologically plausible and useful for research stratification, yet clinical translation must distinguish association from intervention effect. Current antioxidant trial synthesis does not establish improvement in clinical pregnancy, and published guidelines do not support ovarian suppression solely to enhance fertility. The evidence-based next step is to connect standardized molecular measurements with carefully phenotyped clinical cohorts and live-birth outcomes, while maintaining individualized care based on age, symptoms, ovarian reserve, anatomy, partner factors and patient preferences. RWD, causal methods and well-validated AI may complement randomized trials if deployed within reproducible and ethically governed research infrastructure.

DECLARATIONS

Funding: Not reported; to be completed by the author(s). Conflicts of interest: To be declared by the author(s). Ethics approval: Not applicable to this literature review, which involved no original participant data. Data availability: No original dataset was generated. AI assistance and author accountability: To be disclosed in accordance with the target journal's policy. The submitting authors are responsible for verification of all citations and scientific claims.

REFERENCES

1. Becker CM, Bokor A, Heikinheimo O, et al. ESHRE guideline: endometriosis. Hum Reprod Open. 2022;2022(2):hoac009. doi:10.1093/hropen/hoac009.
2. European Society of Human Reproduction and Embryology. Endometriosis guideline [Internet]. 2022. Available from: <https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Endometriosis-guideline>
3. The role of iron in the pathogenesis of endometriosis: a systematic review. Hum Reprod Open. 2023;2023(3):hoad033. Available from: <https://academic.oup.com/hropen/article/2023/3/hoad033/7232232>
4. Ferroptosis and oxidative stress in endometriosis: a systematic review of the literature. 2024. PubMed PMID:38489713. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10939684/>
5. The effects of antioxidant supplementation on pain, oxidative stress markers, and clinical pregnancy rate in women with endometriosis: a systematic review and meta-analysis of randomized controlled trials. 2025. PubMed PMID:41234925. Available from: <https://pubmed.ncbi.nlm.nih.gov/41234925/>
6. US Food and Drug Administration. Data standards for drug and biological product submissions containing real-world data: guidance for industry [Internet]. 2023. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data>
7. European Medicines Agency. Real-world evidence [Internet]. Available from: <https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/real-world-evidence>

8. European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU) [Internet]. Available from: <https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/real-world-evidence/data-analysis-real-world-interrogation-network-darwin-eu>
9. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: WHO; 2021. ISBN:9789240029200.
10. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models [Internet]. 2024. Available from: <https://www.who.int/publications-detail-redirect/9789240084759>
11. European Medicines Agency. A vision for use of real-world evidence in EU medicines regulation [Internet]. 2021. Available from: <https://www.ema.europa.eu/en/news/vision-use-real-world-evidence-eu-medicines-regulation>
12. Davydov AI, Khabarova MB, Chilova RI, Lebedev VA, Shakhlamova MN. Endometriosis and oxidative stress: treatment rationale in endometriosis-associated infertility. Gynecology, Obstetrics and Perinatology. 2023;22(1):69-75. doi:10.20953/1726-1678-2023-1-69-75.
13. World Health Organization. Artificial intelligence for health [Internet]. 2024. Available from: <https://www.who.int/publications/m/item/artificial-intelligence-for-health>
14. European Society of Human Reproduction and Embryology. Guidelines, consensus documents and recommendations [Internet]. Available from: <https://www.eshre.eu/Guidelines-and-Legal>
15. World Health Organization. WHO issues first global report on artificial intelligence in health and six guiding principles [Internet]. 2021. Available from: <https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use/>
16. World Health Organization. Ethics and governance of artificial intelligence for health: executive summary. Geneva: WHO; 2021. ISBN:9789240037403.

Bibliographic note: The 16 entries above are individually identifiable publications or institutional sources. The requested 30–50 independently verified publications were not achieved; additional peer-reviewed studies should be verified and added before

submission. Articles with incomplete metadata must be completed from their original publisher/PubMed record.