

**NON-INVASIVE BIOMARKERS AND ADVANCED IMAGING FOR EARLY  
DETECTION OF ENDOMETRIOSIS: DIAGNOSTIC PERFORMANCE,  
REAL-WORLD EVIDENCE, ARTIFICIAL INTELLIGENCE, AND  
IMPLEMENTATION PRIORITIES**

*Xolboyeva Mashhura Yoqutjon qizi*

*Email – [yoqutjonovnam@gmail.com](mailto:yoqutjonovnam@gmail.com)*

*Termez University of Economics and Service*

*Faculty of Medicine*

**Abstract**

**Background:** Endometriosis is a chronic, heterogeneous inflammatory disease for which delayed recognition and unequal access to specialist imaging remain important clinical problems. The transition from surgically centered diagnosis to an integrated, non-invasive strategy has increased interest in expert transvaginal ultrasonography, magnetic resonance imaging, circulating biomarkers, and saliva-based microRNA signatures. **Objective:** To critically appraise the diagnostic role of these modalities and describe how real-world data, causal inference and artificial intelligence can support safe implementation. **Methods:** A structured critical narrative evidence synthesis was undertaken using identifiable clinical guidelines, regulatory publications, diagnostic-accuracy studies and systematic evidence available through October 2026. This is not a prospectively registered systematic review or a new patient-level analysis; no de novo pooled effect estimates were calculated. **Results:** The 2022 ESHRE guideline endorses ultrasound or MRI during diagnostic work-up but cautions that negative imaging cannot exclude disease, especially superficial peritoneal lesions. A 2025 multicenter external validation of a salivary microRNA signature included 971 symptomatic participants, reported sensitivity of 97.3% and specificity of 94.1%, and requires assessment of transportability, reference-standard variation, and clinical utility before routine adoption. A comparative meta-analysis of rectosigmoid deep endometriosis synthesized 10 studies and 1,604 participants: ultrasound sensitivity/specificity was 0.85/0.92 and MRI 0.83/0.95. These estimates are lesion-site-specific and must not be generalized to all disease phenotypes. Real-world clinical datasets can characterize care pathways, delayed diagnoses, health-service inequality, therapeutic effectiveness, and post-deployment algorithm performance, but do not independently establish diagnostic truth or remove confounding. **Conclusions:** A phenotype-aware, symptom-

led pathway combining expert imaging with appropriately validated adjunctive biomarkers is preferable to reliance on any single test. Responsible use of artificial intelligence requires transparent external validation, calibrated risk estimates, bias audits, prospective clinical impact studies and clear regulatory oversight. Implementation in Uzbekistan should prioritize standardized reporting, structured longitudinal records, workforce development and privacy-preserving diagnostic registries.

**Keywords:** endometriosis; early diagnosis; non-invasive biomarkers; microRNA; saliva; transvaginal ultrasound; magnetic resonance imaging; deep endometriosis; real-world evidence; diagnostic accuracy; machine learning; precision medicine; Uzbekistan.

## 1. Introduction

### Clinical burden and diagnostic uncertainty

Endometriosis cannot be understood as a single imaging phenotype. Ovarian endometrioma, deep infiltrating disease and superficial peritoneal implants exhibit different visibility, biological behavior, symptom associations and risks. The WHO estimated that approximately 10% of reproductive-age women, about 190 million globally, are affected [1]. This estimate describes disease burden, not the prevalence in a referred diagnostic cohort. Pain severity is not a reliable surrogate for lesion extent; alternatively, severe organ involvement may be present in women whose symptoms are non-specific. Delayed recognition can arise when dysmenorrhea is normalized, when infertility is assessed without structured pain history, or when a normal routine scan is incorrectly interpreted as evidence against the disease. These distinctions matter because clinically meaningful early detection should accelerate appropriate care rather than simply increase positive test counts.

### From laparoscopy to non-invasive clinical pathways

Historical diagnostic paradigms privileged laparoscopic visualization and histology. Contemporary ESHRE guidance recognizes a greater role for high-quality imaging and empirical management, while preserving laparoscopy for selected patients with negative imaging or unsuccessful or unsuitable empirical treatment [2,3]. The shift is not a declaration that imaging detects all endometriosis. Rather, it reflects the balance between diagnostic information, surgical exposure, resource availability and patient

preference. In practice, the clinician must ask which phenotype is being sought, at what anatomical location, by which operator, and for which decision. A scan that adequately assesses an ovarian endometrioma may still fail to document posterior compartment mobility or superficial peritoneal lesions.

### **RWD and RWE: concepts and distinctions**

Real-world data (RWD) are routinely collected observations arising from clinical practice, including electronic health records, pharmacy dispensing, laboratory systems, hospital registries, administrative claims and patient-generated digital information. Real-world evidence (RWE) is the clinical evidence about the use, benefits or risks of an intervention derived from analysis of such data. It is therefore a category error to equate the existence of a large database with strong evidence. In endometriosis, routinely captured data can measure time from first symptom-coded encounter to diagnosis, practice variation in imaging referral, persistence with hormonal treatment, postoperative re-intervention and inequalities in access. However, symptom documentation is often incomplete and the coded date of diagnosis can differ substantially from the biological onset of the disease [13-15].

### **RCTs and observational evidence**

Randomized controlled trials (RCTs) seek causal estimates by allocating interventions at random under defined protocols, thus mitigating measured and unmeasured baseline confounding in expectation. RWE studies observe care decisions shaped by symptoms, previous treatment, patient choice and clinician preference. They can offer greater generalizability and longer follow-up, but are vulnerable to confounding by indication, differential ascertainment and immortal-time bias. These approaches are complementary, not substitutes: RCTs may evaluate whether a diagnostic strategy or drug changes patient-relevant outcomes, while RWD may illuminate adoption, off-label patterns, access barriers and effectiveness across heterogeneous clinical settings. Diagnostic-accuracy studies are a third evidentiary category; sensitivity and specificity alone do not show that earlier testing improves pain, fertility or quality of life.

### **Efficacy versus effectiveness**

Treatment efficacy refers to a therapy's effect in controlled experimental conditions, while effectiveness describes its performance in routine care. For endometriosis, trial

eligibility may exclude complicated comorbidity or concurrent hormonal regimens that are common in practice. Treatment effectiveness may depend on adherence, adverse effects, affordability and clinician follow-up. The same distinction applies to diagnostic tools: technical discrimination in a carefully selected validation cohort is not equivalent to improved decisions in a primary-care population. A technically precise classifier may have low positive predictive value when deployed in a lower-prevalence population, producing unnecessary referrals and anxiety.

## **International policy and artificial intelligence**

FDA guidance provides a framework for using electronic records and claims to support regulatory inferences, emphasizing data relevance, reliability and transparency [13-15]. WHO guidance identifies autonomy, safety, explainability, accountability, equity and sustainability as central to responsible health AI [16-18]. FDA guidance on AI-enabled medical-device software adopts a lifecycle perspective, including predetermined change control plans [19-22]. These general frameworks are relevant to imaging algorithms and biomarker interpretation; they must not be misread as approval of any particular endometriosis diagnostic product. The European regulatory emphasis on RWE and networked data generation can inform governance, but local legal compliance and independent validation remain essential.

## **Purpose and scope**

The objective of this review is to evaluate the relative and complementary roles of imaging, biomarkers and AI in recognizing endometriosis earlier, while articulating a methodologically sound RWD/RWE agenda for longitudinal evaluation. The review also examines pharmacovigilance and personalized treatment as downstream uses of the same information infrastructure. Examples from diabetes, cardiology, oncology, infection and neurology are presented only as methodological transfer opportunities; evidence from those specialties cannot be used as evidence that an endometriosis tool works.

## **2. Materials and Methods**

### **Review design and reporting status**

This manuscript is a structured critical narrative review integrating independently identifiable guidelines, regulatory documents, consensus statements, and selected

primary diagnostic studies. It does not claim a completed PRISMA systematic review, a registered protocol, exhaustive database retrieval, formal meta-analysis or patient recruitment. Sources were located using targeted public bibliographic and institutional searches for material available by 16 October 2026. PubMed-indexed reports and official WHO, ESHRE and FDA publications provide the core verifiable basis. Scopus, Web of Science, Embase, Cochrane Library, Nature Medicine, The Lancet, NEJM and JAMA are discussed as appropriate bibliographic and journal sources, but independent comprehensive searches of all subscription platforms were not performed and must not be asserted as completed.

### **Eligibility framework**

Priority was given to adult or adolescent populations with suspected or confirmed endometriosis; studies addressing transvaginal ultrasound, MRI, biochemical or molecular biomarkers; prospective diagnostic validation; evidence synthesis; guidelines; and transparent regulatory policy. Preferred publication dates were 2020–2026, with seminal consensus work retained when essential for protocol definitions. Exclusion from quantitative comparison applied to non-endometriosis populations, tests evaluated only in animals, convenience case–control studies without transportability assessment, unverified promotional statistics, inaccessible unconfirmed citations and duplicate publications of identical consensus work. Narrative context can still describe exploratory findings as hypotheses rather than validated clinical tests.

### **Evidence extraction and methodological appraisal**

Data elements included population spectrum, recruitment setting, index-test protocol, operator training, sample and handling conditions, reference standard, blinding, lesion phenotype, disease prevalence, sensitivity, specificity, uncertainty intervals, external validation, conflicts of interest and implementation status. QUADAS-2 domains inform the appraisal of patient selection, index test, reference standard and flow/timing. Risk of bias is conceptually distinguished from statistical imprecision. Because there was no exhaustive search and study-level re-extraction, numerical pooling, publication-bias tests and grading certainty with GRADE would give a false appearance of quantitative completeness.

### **Causal inference and target trial emulation**

A target-trial emulation specifies eligibility, time zero, assigned strategies, follow-up, outcome, estimand, analysis and censoring before observational data analysis. An endometriosis example would compare immediate expert ultrasound referral against usual referral among symptomatic patients who meet the same baseline criteria. The outcome might be time to appropriate specialist-directed management, not simply time to a coded diagnosis. Time zero must be identical for both arms; treatment switching, indication severity, and availability of ultrasound expertise require explicit management. The analysis remains vulnerable to unmeasured confounding and inconsistent documentation even when design is careful.

## **Propensity scores and inverse probability weighting**

Propensity score matching attempts to balance measured pre-exposure variables between comparison groups; inverse probability weighting reweights observations by their probabilities of receiving the observed exposure. Candidate confounders include age, symptom burden, prior imaging, fertility goals, socioeconomic position, referral channel and comorbidity. Diagnostics include overlap plots, stabilized weights and standardized mean differences, not merely an apparently favorable p value. Extreme weights indicate positivity problems, while adjustment for post-exposure variables can bias effects. Neither approach repairs unknown confounders or poorly defined exposure dates.

## **Survival and longitudinal methods**

Kaplan–Meier curves summarize event-free experience and require honest censoring assumptions; Cox proportional hazards regression estimates conditional hazard ratios if proportional hazards and model assumptions are acceptable. In endometriosis datasets, follow-up can evaluate delay to referral, surgery, symptom recurrence or treatment discontinuation, with competing events and informative loss to follow-up addressed where relevant. Diagnostic sensitivity is not a survival endpoint. Recurrent clinical encounters and symptom fluctuation may call for multistate models, recurrent-event models or mixed effects rather than a simplistic first-event Cox model.

## **Machine learning, deep learning and NLP**

Logistic regression and gradient-boosting classifiers offer benchmark prediction approaches; convolutional or transformer image models may process sonograms or

MRI, while natural language processing (NLP) can structure symptom narratives and radiology reports. Models should be developed using patient-level separation, nested resampling when appropriate, representative external sites, calibration assessment and decision-curve or clinical utility analyses. Large language models may support documentation extraction but can hallucinate diagnoses, silently omit negation or propagate dataset inequity. Reinforcement learning for adaptive referral pathways remains a research concept rather than established endometriosis care.

### Diagnostic metrics and synthesis rules

Sensitivity and specificity refer to classification relative to an explicit reference standard; likelihood ratios and predictive values help translate them to practice. Predictive values depend on pretest probability. Results by bowel segment, ovarian site or superficial disease must not be combined into an undifferentiated global accuracy score. Confidence intervals are reproduced only when explicitly supported by a verified publication. No new diagnostic pooled estimate, hazard ratio, matched-cohort treatment effect or AI performance metric was calculated for this article.

## 3. Results: Critical Synthesis of the Available Evidence

### 3.1 Diagnostic landscape and benchmark findings

Across the reviewed guidance and clinical studies, the decisive distinction is phenotypic visibility. ESHRE recommends ultrasound and/or MRI for suspected endometriosis, while explicitly noting that negative findings do not exclude superficial peritoneal disease [2,24,25]. The 2024 international imaging consensus supports harmonized lesion mapping and classification by trained teams [9,10]. The 2025 prospective diagnostic-accuracy study comparing ultrasound with surgery documented particularly poor sensitivity for some superficial disease sites, notwithstanding strong performance for endometriomas and deep lesions [11]. Consequently, an early-detection pathway should separate detection of visible structural disease from management of persisting symptoms when structural imaging is negative.

| Evidence     | Population design | / Published finding |               | Interpretation |                                      |
|--------------|-------------------|---------------------|---------------|----------------|--------------------------------------|
| WHO 2025 [1] | Global estimate   | burden              | ~10%; million | ~190           | Burden estimate, not test prevalence |

|                                               |                                                    | reproductive-age women                                                      |                                                                     |
|-----------------------------------------------|----------------------------------------------------|-----------------------------------------------------------------------------|---------------------------------------------------------------------|
| <b>Saliva miRNA</b>                           | <b>Multicenter</b>                                 | <b>Sensitivity 97.3%; specificity 94.1%; 95% CI 96.4–98.0 and 91.0–96.4</b> | <b>Promising selected symptomatic cohort; needs impact evidence</b> |
| <b>2025 [4]</b>                               | <b>external validation; n=971</b>                  |                                                                             |                                                                     |
| <b>Rectosigmoid imaging meta-analysis [5]</b> | <b>10 studies; n=1,604</b>                         | <b>TVUS Se 0.85 / Sp 0.92; MRI Se 0.83 / Sp 0.95</b>                        | <b>Site-specific pooled estimates; not superficial disease</b>      |
| <b>IDEA-based TVUS 2025 [11]</b>              | <b>Prospective surgical-reference study; n=125</b> | <b>Superficial-site sensitivity 4.0–43.5%</b>                               | <b>Negative examination cannot exclude superficial disease</b>      |

### 3.2 Expert transvaginal ultrasound, dynamic mapping and adolescent adaptations

Transvaginal ultrasonography is frequently the first-line imaging technique because it combines availability, dynamic examination, anatomic resolution and comparatively modest cost. Expert examination is not equivalent to a routine generic pelvic scan. IDEA-based practice addresses the uterus and adnexa, the sliding sign, tenderness-guided examination, soft markers and structured anterior/posterior compartment mapping [6,11,12]. A technically adequate scan also documents which compartments could not be assessed; unspecified “pelvis normal” language is a poor substitute. When transvaginal examination is inappropriate or unacceptable, transabdominal, transperineal and selected transrectal modalities or MRI may be considered, particularly for younger patients. Clinical consent and patient preference remain essential.

An important innovation in 2025 was the IDEA addendum focusing on superficial disease. It standardizes emerging sonographic observations while implicitly acknowledging that further validation is needed [7]. Earlier diagnosis should not be equated with finding all implants in every patient. For deep lesions, technical training and protocol harmonization may matter more immediately than introduction of a black-

box classifier. Imaging also serves presurgical planning and anticipates the need for colorectal or urological collaboration, a different endpoint from diagnostic triage.

### **3.3 Magnetic resonance imaging and complementary lesion characterization**

MRI is especially useful in defining complex anatomy, deep compartment extent, lateral pelvic involvement and ambiguous findings when performed with an optimized pelvic protocol and expert interpretation. MRI has an advantage in broad field of view and multiplanar mapping, whereas dynamic ultrasound provides real-time mobilization information. Such complementary strengths do not imply that MRI should replace high-quality ultrasound for all patients. Imaging choice depends on suspected phenotype, local expertise, contraindications, clinical objectives and cost. Rectosigmoid meta-analysis results are close in pooled performance, with ultrasound sensitivity/specificity 0.85/0.92 and MRI 0.83/0.95 across the reported evidence [5].

An MRI report that lacks lesion-compartment mapping may fail to support multidisciplinary planning even if it correctly identifies endometriosis. Conversely, high-resolution images may create false reassurance if superficial lesions remain below practical detection. Because pathology is a selective reference standard and surgery is more likely in complicated patients, published diagnostic accuracy can be affected by verification and referral bias. Methodologically credible MRI research should disclose scan sequences, reader experience, blinded interpretation, target lesion sites and reference-standard timing.

### **3.4 Blood-based biomarkers, CA-125 and inflammatory panels**

Serum CA-125 is biologically plausible because peritoneal inflammation can elevate circulating antigen concentrations. However, physiological cycle variation, ovarian disease and numerous non-endometriosis conditions compromise specificity and sensitivity. ESHRE explicitly recommends against using tissue, blood, menstrual or uterine-fluid biomarkers alone to establish the diagnosis [2,24]. Multiplex inflammatory panels incorporating cytokines, angiogenesis mediators or immune-cell signatures have not yet overcome the practical obstacles of between-laboratory reproducibility, sample timing and disease-spectrum bias. A statistical difference in mean biomarker concentration between surgical cases and healthy controls does not by itself yield a clinically safe test.

The critical issue is not whether a marker correlates with the presence of lesions but whether a prespecified threshold performs in the population where clinicians intend to order the assay. Prospective all-comers studies must include symptomatic alternative diagnoses and mild disease, not just advanced cases. Measurement error, hormonal exposure, menstrual-cycle timing, coexisting adenomyosis, obesity and inflammation require explicit characterization. When a biomarker is combined with symptoms and ultrasound, the incremental benefit should be measured against the clinical baseline model rather than against chance.

### **3.5 Salivary microRNA and other multi-omic signatures**

Circulating and salivary microRNAs may capture regulatory pathways involving inflammation, apoptosis and epithelial-mesenchymal biology. Their non-invasive sampling is attractive for young patients and longitudinal reassessment, but sequencing pipelines can be sensitive to collection containers, storage time, hemolysis or oral inflammation. In the 2025 ENDOmiRNA external validation study, 971 symptomatic patients were evaluated, with endometriosis prevalence of 77%; reported sensitivity was 97.3% (95% CI 96.4–98.0) and specificity 94.1% (95% CI 91.0–96.4) [4]. The reported positive and negative predictive values are conditional on the high-prevalence study setting and therefore cannot be directly exported to population screening.

The external validation is stronger evidence than discovery-stage biomarker comparisons, yet it does not settle clinical utility. The study used imaging and/or surgery to determine disease status, creating potential heterogeneity in reference standard; the way controls were established and the prevalence spectrum require careful interpretation. Funding by the test developer, documented in the publication, should be transparently considered alongside scientific merit [4]. Future independent, geographically diverse validation should prioritize calibration, missed superficial lesions, diagnostic delay, patient-centered outcomes and cost-effectiveness.

### **3.6 Urine, menstrual effluent, proteomics and metabolomics**

Urinary peptide, extracellular vesicle, metabolite and cell-free RNA candidates may reflect pelvic inflammation or systemic response. Menstrual effluent is conceptually closer to the affected tissue environment but is harder to standardize for cycle day, sample contamination and hormonal treatment. Across these modalities, the central translational challenge is repeated demonstration of stable individual-level

classification independent of sample batch and case–control selection. A high discovery AUC in a small cohort can fall substantially in routine symptomatic referrals.

Proteomic and metabolomic pipelines are particularly susceptible to high-dimensional overfitting. Quality-control samples, pre-analytical standard operating procedures, lockbox validation, replicate measurements and prespecified assay failure rules must precede adoption. Multi-omics integration may identify biologically coherent subtypes, yet adding omics layers increases costs and missingness. An ideal assay should define whether it diagnoses the disease, characterizes lesions, predicts progression, or forecasts response; these are distinct intended-use claims.

### **3.7 Combining symptoms, imaging and biomarkers**

A clinically useful strategy can begin with a structured symptom history, examination when acceptable, and an expert imaging referral targeted to the suspected anatomical pattern. Adjunctive biomarkers are more defensible when they resolve a clearly defined uncertainty, such as prioritizing further referral among symptomatic people whose initial imaging is negative. A risk model should not automatically substitute a positive biomarker for specialist assessment, and it should not use negative biomarker results to deny reassessment when severe cyclical pain persists.

Multimodal models should be evaluated for calibration, sensitivity at clinically defensible thresholds, net benefit and the consequences of false reassurance. Pretest probability varies between adolescents in primary care, infertility clinics and tertiary referral centers. Thresholds appropriate for a high-prevalence specialist population may overload imaging services if transferred unchanged into general practice. Explicit “do not miss” pathways are especially important for bowel, urinary tract and obstructive disease.

### **3.8 Electronic medical records and electronic prescribing as RWD sources**

Longitudinal EHRs can connect clinical history, coded symptoms, imaging interpretations, analgesic use, hormonal therapy, surgery and functional outcomes. Well-designed data dictionaries are essential because endometriosis can be coded as suspected, probable, surgically confirmed or historical; treating these as identical creates misclassification. Free-text records may contain the most relevant features, including cyclical dyschezia or school absenteeism, but their structure and language

vary. Electronic prescribing can capture treatment initiation and switching, although prescription is not synonymous with dispensing or adherence.

In endometriosis, real-world “time to diagnosis” should ideally begin at the first documented relevant symptom rather than the first specialist visit. Left-censoring, episodic attendance and migratory care must be examined. De-identified linkage across EHRs and pharmacies can distinguish intolerance-related discontinuation from supply limitations; without explicit causal design, however, observed treatment persistence must not be called treatment efficacy. FDA guidance on EHR/claims emphasizes fitness for purpose and traceability [14].

### **3.9 Laboratory databases, registries and insurance claims**

Laboratory databases can support exploration of inflammatory and hematological trajectories and verify assay procedures, but specimen handling and indication bias complicate analysis. Disease registries can anchor standardized phenotyping, surgical mapping and longitudinal outcomes across facilities. Claims datasets offer scale for tracking healthcare utilization, imaging, medication and reoperations, while commonly lacking detailed symptom severity or lesion maps. These datasets complement rather than replace clinical examination and reference-standard assessments.

A reliable registry should record the date and basis of diagnosis, suspected phenotype, standardized imaging location, treatment exposure, fertility intentions, and validated patient-reported outcomes. Linkage rules should be audited for duplicate records, changes of surname, referral transfers and missing encounters. The purpose of a registry is not simply to collect more records, but to create reproducible information that supports clinically interpretable comparisons.

### **3.10 Wearables, mobile applications and patient-generated information**

Cycle-tracking applications, sleep/activity monitors and symptom diaries may capture the temporal structure of pain, fatigue and functional impairment that is poorly represented in episodic clinic visits. These signals can improve communication and reveal delayed referral or cyclical symptom patterns. They cannot diagnose histological endometriosis by themselves, and symptom variability or menstrual-cycle association is not specific. App data are prone to self-selection, differential access, device measurement error and loss to follow-up.

User-controlled export, consent, secure storage, transparent data retention and usability for diverse literacy levels are vital. An application that promises screening without external validation may increase alarm or falsely reassure. Clinical value should be assessed as improvement in referral appropriateness, symptom communication or time to suitable treatment, not merely engagement metrics or classifier accuracy.

### **3.11 Machine learning and deep learning for imaging**

Machine learning can synthesize structured symptom profiles, morphological imaging descriptors and laboratory covariates. Deep learning can assist lesion segmentation, image-quality assessment and identification of patterns that are difficult to formalize manually. However, labeled training images often come from expert centers and devices with distinct acquisition presets. Models may learn scanner artifacts, annotation conventions or referral prevalence instead of transferable disease features. A model trained on patients scheduled for surgery must not be presented as a validated community screening tool.

Essential safeguards are patient-level train-test separation, site-held-out validation, prospective silent deployment, calibration testing, subgroup audits and reporting of failure modes. An interpretable baseline is needed to show whether AI improves over expert ultrasonography. No verified evidence reviewed here justifies claiming that an AI algorithm independently exceeds specialist imaging performance for all endometriosis phenotypes. Explainable-AI heatmaps can reveal areas of model attention but do not prove causal pathophysiology.

### **3.12 NLP, LLMs and adaptive clinical decision support**

NLP can transform free-text notes into time-stamped symptom features, detect negation and identify prior inconclusive scans. LLM systems can support referral-letter drafting, report harmonization and patient education, but generative output is not a reference standard. Clinical records frequently contain copied forward diagnoses and shorthand that may confuse automated extraction. Any AI-assisted documentation service requires human review of clinical facts, version control, representative evaluation datasets and incident reporting.

Reinforcement learning is sometimes proposed to optimize referral sequences or testing thresholds. Its use in this domain is preliminary: reward functions may trade shorter coded diagnosis time against avoidable investigations or unequal access, and

policy learning from historical decisions may amplify bias. A safer initial use of CDSS is guideline-concordant prompts that flag persistent cyclical symptoms and incomplete imaging maps, while preserving clinician accountability [16-22].

### **3.13 Treatment-response prediction and cross-specialty lessons**

Prediction of pharmacological response could eventually help select hormonal therapies and anticipate discontinuation, but endometriosis treatment choice must incorporate fertility intentions, contraindications, prior adverse effects and patient preference. Transferable examples from diabetes include individualized glycemic outcomes and confounding by baseline disease severity; cardiovascular studies illustrate longitudinal safety and competing risks; oncology emphasizes biomarker-treatment interactions and external validation; infectious diseases require time-varying exposure accounting; neurological disorders highlight symptom heterogeneity and patient-reported endpoints. These are analytic analogies, not claimed disease-specific trial findings.

Machine-learning risk scores for pain persistence should be judged by incremental benefit over transparent clinical models and their impact on decisions. Diagnostic and prognostic models must not be conflated. A biological biomarker associated with the presence of endometriosis may perform poorly for predicting progestin response; conversely, a drug-response predictor need not detect lesions. Prospective treatment-strategy comparisons remain necessary before decision systems guide therapy.

### **3.14 Pharmacovigilance and adverse drug reactions**

Medication safety surveillance can identify suspected adverse drug reactions associated with analgesics, hormonal agents and other therapies used in patients with endometriosis. Spontaneous reporting systems support signal generation but cannot independently determine incidence or causality because reporting is incomplete, stimulated by publicity and affected by denominator uncertainty. EHR-linked cohorts can assess pre-specified outcomes using active comparators, while accounting for contraindications and baseline risk. Pharmacovigilance analytics should separate new safety signals from established adverse-effect profiles and reflect available product information.

Drug-safety signal detection may use disproportionality analyses or sequence-based methods; these are hypotheses requiring clinical review. In an endometriosis registry,

clinically important follow-up includes treatment discontinuation due to adverse effects and effects on daily function. AI can accelerate narrative case processing but should not silently substitute code-based classification for clinician causality assessment. Reproductive and pregnancy status heighten the need for careful exposure definition.

### **3.15 Precision medicine, pharmacogenomics and biological heterogeneity**

Precision medicine seeks to align investigation and therapy with a clinically meaningful disease subtype or treatment-response profile. Endometriosis varies by anatomic site, inflammatory activity, age, hormonal milieu, central sensitization and coexistence with adenomyosis or pelvic-floor disorders. A biomarker should be labeled diagnostic, prognostic, predictive or pharmacodynamic according to its validated purpose. Pharmacogenetics describes specific inherited variant effects; pharmacogenomics more broadly examines genomic contributions, often in combination with environmental and biological context.

Although omics research is expanding, routine biomarker-guided drug selection in endometriosis lacks a broadly established, clinically validated pathway. A precision framework must therefore integrate patient goals and structured phenotype description before expensive molecular panels. Evidence of analytical validity, clinical validity and demonstrated utility is required in sequence. Claims of personalized care must be tied to a measured improvement in patient-relevant outcomes, not merely to the use of sequencing.

### **3.16 Practical diagnostic pathway and uncertainty management**

At first contact, clinicians should recognize persistent or progressive severe menstrual pain, chronic pelvic pain, dyspareunia, bowel or urinary symptoms associated with cycles, fatigue and infertility [2,23]. A structured history and focused examination should guide expert imaging; MRI can supplement uncertain or complex anatomical mapping. When imaging is negative, the patient should not be told that endometriosis has been excluded. Decisions about empirical hormonal therapy, reassessment and possible laparoscopy should reflect preferences, contraindications, symptom burden and the differential diagnosis.

Adjunctive salivary or molecular testing should currently be assessed through local regulatory status, independent performance evidence and the availability of appropriate

counseling. A “positive” or “negative” biomarker cannot replace clinical judgment in an unvalidated setting. Implementation metrics should include the proportion receiving structured imaging, time to appropriate treatment, burden of unnecessary procedures, missed complicated lesions and patient-reported quality of life.

### 3.17 Data systems: comparative value for endometriosis RWE

| Source              | Useful endometriosis variable     | Major limitation            | Data-quality safeguard          |
|---------------------|-----------------------------------|-----------------------------|---------------------------------|
| EHR                 | Symptom onset, imaging, treatment | Missing history; miscoding  | Structured phenotypes and audit |
| E-prescription      | Initiation and switching          | Prescription adherence      | Link with dispensing            |
| Laboratory          | Assay results and trends          | Preanalytic variation       | Timestamp / batch metadata      |
| National registry   | Longitudinal lesion mapping       | Coverage and selection bias | Common data model               |
| Pharmacovigilance   | Suspected adverse reaction        | Under-reporting             | Case validation                 |
| Insurance claims    | Resource use and surgery          | Sparse clinical detail      | Clinical linkage                |
| Wearables           | Activity and sleep trends         | Device selection bias and   | Device consent QA;              |
| Mobile applications | Cycle-related symptoms            | Self-report and attrition   | Validated instruments           |

### 3.18 Statistical designs and interpretation of endometriosis RWE

#### Comparative effectiveness target trial

A practical emulation would assess specialist imaging referral within a defined interval versus a usual-care pathway among comparable symptomatic patients. Define eligibility before treatment/referral choice, a single time zero, follow-up windows and safety/utility endpoints. Confounding by pain severity and fertility goals is a major

threat. Matching or weighting can address measured differences if positivity and measurement assumptions hold.

### **Time-to-event models**

Kaplan–Meier curves can describe referral or recurrence trajectories with reported censoring distributions. Cox regression offers a conditional hazard ratio if proportional-hazards assumptions are examined. Competing events, such as pregnancy or hysterectomy, may change interpretation; death is not the only competing event in longitudinal reproductive care. Hazard ratios should not be invented from aggregate diagnostic summaries.

### **Prediction validation**

Machine-learning work requires separate tasks for diagnosis, progression and treatment response. Discrimination alone is insufficient: calibration curves, net-benefit analysis and external validation are needed. Non-independent imaging frames from the same patient or physician lead to data leakage. Threshold performance should be analyzed against the cost of false reassurance and unnecessary escalation.

### **Missing-data mechanisms**

Missing completely at random, missing at random and missing not at random carry different implications. Multiple imputation may be useful for incomplete measured confounders under assumptions, but cannot reconstruct absent symptom episodes in routinely coded records. Sensitivity analyses should explore plausible deviations, and complete-case analyses should not be accepted uncritically.

### **3.19 Diagnostic test accuracy, prevalence and decision thresholds**

The interpretation of a diagnostic test depends on the intended decision, not solely on its numerical sensitivity and specificity. Positive predictive value represents the proportion of positive test results that correspond to disease within the tested population, and negative predictive value represents the analogous proportion of true absence among negative results. Both quantities change with prevalence even if sensitivity and specificity remain constant. Thus, a salivary signature validated in a specialist population with 77% prevalence cannot automatically be expected to produce the same predictive values in adolescents presenting with nonspecific dysmenorrhea at

a primary-care clinic [4]. A transparent economic and clinical model should vary pretest probabilities, referral capacity, test failure and access costs. Decision thresholds must also balance the asymmetric harms of a missed bowel lesion and an unnecessary specialist referral. A rule-out claim demands particularly convincing sensitivity in the intended low-prevalence environment and independent evidence about superficial lesions. Reporting one summary accuracy number risks concealing these considerations.

Likelihood ratios provide a more transportable mathematical link between a result and pretest odds, yet their stability also presupposes comparable disease spectrum and verification methods. They are not biological constants. For serial pathways, the performance of each stage depends on the selection induced by previous stages. If imaging-negative individuals alone receive a molecular test, the assay is operating in a conditional subgroup whose disease spectrum differs from the original biomarker validation study. Diagnostic studies should therefore reproduce the proposed clinical sequence in the same cohort. When comparing combinations of clinical history, imaging and molecular testing, investigators should report incremental discrimination alongside calibration, reclassification consequences and decision-curve net benefit, avoiding a misleading inference that a modest AUC improvement necessarily produces patient benefit.

### **3.20 Reference standards, disease verification and latent misclassification**

Endometriosis creates a difficult reference-standard problem. Laparoscopy with appropriately interpreted histology is often treated as definitive when surgery is undertaken, but patients with less severe symptoms or a reassuring scan are less likely to undergo surgery. This introduces partial verification bias because the reference standard is applied according to the index test or clinical suspicion. Meanwhile, expert imaging can establish clinically actionable ovarian or deep disease without surgery, although it is not equally sensitive for superficial implants. Combining imaging and surgery into a composite reference may improve practical inclusiveness yet permits dependence between the index and reference tests. Consequently, a diagnostic validation study must state whether imaging itself contributed to case classification and whether all controls had comparable disease verification. The ENDOMiRNA study reported surgical evaluation of controls, an important design detail that warrants careful consideration together with its other reference-standard choices [4].

Histological negative findings do not necessarily refute endometriosis when sampling is limited, and surgical inspection also has operator limitations. Accordingly, sensitivity estimates should not be treated as precise measures of an immutable latent truth when the reference procedure is imperfect. Protocols can specify independent adjudication, blinded imaging review, standardized operative documentation and follow-up to reduce uncertainty. Advanced methods such as latent-class models can explore imperfect reference standards, but these approaches depend on strong and often unverifiable assumptions regarding conditional independence and population composition. A rigorous paper should prioritize honest uncertainty about disease verification over the appearance of mathematical certainty.

### **3.21 Imaging quality assurance and reporting comparability**

The diagnostic performance of imaging systems is partly produced by institutions rather than devices. Ultrasound quality depends on examination time, equipment, transducer choice, operator experience and the ability to perform dynamic tests. MRI performance depends on protocol standardization, field strength, sequence choice and expert interpretation. A multicenter implementation program should therefore document not only machine manufacturer and nominal modality but also the proportion of examinations conforming to protocol, competence levels of operators, structured report completeness and rates of technically limited studies. The IDEA statement and subsequent addenda provide a starting vocabulary for consistent lesion location and compartment mapping [6-8]. A standardized structured report can facilitate longitudinal comparison and multidisciplinary planning, although it cannot compensate for an inadequately acquired examination.

Quality assurance is essential when comparing clinics using RWD. Referral centers may receive a higher concentration of deep disease and employ more experienced sonographers; naive comparisons of positive-scan rates would therefore confound disease spectrum with performance. A defensible benchmarking approach adjusts for referral case mix and assesses inter-reader agreement, blinded re-reading and clinically relevant discrepancies. Report-level auditing can identify missing sliding-sign documentation, incomplete bowel or urinary assessment, and unrecognized uncertainty. These process metrics may provide more immediate gains in real-world diagnostic reliability than replacing reporting staff with automated image classification.

### **3.22 Data harmonization, interoperability and reproducibility**

RWD studies across hospitals require stable semantic definitions of symptoms, interventions, imaging compartments, and outcomes. A diagnosis date may represent the first provisional code, first expert scan, histological confirmation or time of discharge; pooling these dates without a shared rule invalidates comparisons. Common data models can harmonize high-level concepts, but mapping free-text phrases such as cyclical bowel pain or focal posterior tenderness remains challenging. Provenance metadata should distinguish physician-entered evidence, patient-reported history, machine-generated interpretation and retrospective coding. Missing timestamps and copied-forward diagnosis labels demand particular attention. The FDA data-standards guidance offers general regulatory rationale for standardized electronic submissions and traceable source data [15].

A reproducible study should maintain versioned codebooks, audit trails, site-specific extraction scripts and prespecified transformation rules. Phenotype labels should be sufficiently granular to avoid conflating endometrioma with superficial peritoneal findings; the ability to encode simultaneous phenotypes is useful. Multi-institutional databases must resolve duplicated visits, transfers, pregnancy intervals, hormonal treatment changes and diverse report languages. In Uzbekistan, these practical interoperability challenges make development of a minimum common dataset more feasible as an initial goal than a nationwide machine-learning prediction service.

### **3.23 Confounding, selection and fairness in longitudinal pathway analyses**

Observed treatment choice and specialist referral are not random events. Severe pain, previous infertility treatment, health literacy, insurance or out-of-pocket capacity, rural residence and clinician judgment can influence who receives MRI or specialist ultrasound. These same variables may influence subsequent treatment and measured time to diagnosis. Conventional adjustment, propensity matching and inverse probability weighting can balance observed baseline characteristics but cannot correct for unknown preferences or undetected disease severity. The positivity assumption can fail when some patients have no realistic access to specialist scanning, rendering weighted comparisons unstable or conceptually inappropriate. A target-trial formulation should identify the actual clinical choice available to each eligible person, explicitly restrict comparisons to populations with reasonable treatment overlap, and distinguish intention-to-refer from completed imaging.

Algorithmic fairness further complicates implementation. A classifier with acceptable pooled sensitivity can still perform poorly in subgroups if their symptoms or imaging data were underrepresented in training. Device type, age, rural referral patterns and disease phenotype should be evaluated prospectively as prespecified strata. Fairness cannot be resolved only through demographic balance in the training data; it also concerns outcomes of error, clinical recourse and differential access to confirmatory care. A false negative that delays specialist review can have more serious consequences for patients with progressive urinary obstruction than for a patient with stable uncomplicated symptoms. Therefore, subgroup calibration and decision consequences should accompany subgroup discrimination metrics.

### **3.24 Missing data, privacy and federated learning**

Missing clinical data may be a symptom of the health system rather than random technical failure. A symptom can be absent from a record because it was not asked, was not disclosed, or was considered unrelated to the visit. This mechanism can correlate with clinician expertise and patient background. A statistical imputation model may offer sensible estimates under a missing-at-random assumption, but cannot reliably repair systematic non-recognition of disease. Investigators should provide a variable-level missingness table, compare data completeness across clinical settings and perform sensitivity analyses for plausible departures from assumptions. Complete-case models preferentially select patients with comprehensive specialist documentation, potentially overestimating performance in ordinary care.

Sensitive reproductive-health information warrants proportional governance. Federated learning may allow selected algorithms to train on records held at participating institutions, reducing transfer of raw data, but does not itself solve privacy, security, representativeness or governance problems. Model updates can leak information if improperly designed; differential privacy, secure aggregation and access control require evaluation in context. Consent, lawfully defined access, data minimization and independent oversight remain prerequisites. Longitudinal records must also allow auditing of erroneous linkage or patient-requested corrections without compromising analytic traceability.

### **3.25 Prospective implementation and health-economic evaluation**

A diagnostic innovation should ultimately be judged through its effect on clinical pathways and patient-relevant outcomes. A pragmatic prospective study could compare usual symptom-led care with a strengthened pathway incorporating an expert imaging template and structured referral prompts. Primary outcomes might include time to appropriate diagnostic assessment, patient-reported pain-related functional interference, or proportion of clinically consequential disease identified within a prespecified interval. Secondary outcomes could include unhelpful imaging, unnecessary surgery, treatment initiation, delayed complicated disease and clinician workload. Importantly, shorter time to a positive label is not necessarily beneficial if it results from overdiagnosis or excessive testing. Studies should measure harms and costs alongside perceived benefit.

Health-economic models should itemize the costs of imaging training, MRI slots, molecular assays, laboratory infrastructure, confirmatory consultations and future surveillance. Assumptions about test prevalence, diagnostic accuracy and pathway uptake should be explicit and varied in sensitivity analysis. In a resource-limited setting, improved imaging quality may outperform universal molecular screening despite less striking headline technology claims. Independent local cost and implementation evidence is required before proposing national procurement. Research protocols should specify how incidental ovarian masses, negative imaging with persistent symptoms and urgent organ-involvement findings are handled so that randomized or observational pathway comparison does not compromise appropriate care.

## 4. Discussion

### **Synthesis with international guidance**

The strongest actionable message is a layered interpretation rather than a competition between molecular diagnostics and imaging. Established imaging recommendations retain immediate clinical applicability for endometriomas and deep disease [2,9,10], whereas the exceptionally high reported salivary-miRNA performance is promising but context-dependent [4]. These two evidence types are not directly comparable without matched cohorts, consistent reference standards and a clear clinical decision. The 2025 superficial-imaging consensus addendum [7] also reinforces how technology and examination technique evolve: a negative scan should prompt context-sensitive reassessment, not diagnostic closure.

### **Bias, applicability and spectrum effects**

Spectrum bias arises when validation relies on high-prevalence tertiary populations or surgically selected cases. Verification bias arises when only some women undergo definitive operative investigation; reference-standard heterogeneity can further distort estimates. Operator experience is a causal component of real-world imaging performance. Biomarker signals may reflect cycle phase or coexisting inflammatory disorders. Diagnostic development studies should prospectively enroll consecutive symptomatic patients and compare the new test with explicit current-care pathways rather than highly selected healthy controls.

### **Explainability and clinical oversight**

Explainable AI can improve debugging, particularly if saliency maps highlight irrelevant image margins or metadata, but an explanation is not a guarantee of clinical correctness. A clinical decision-support system should disclose intended population, limitations, uncertainty and audit status. The WHO guidance prioritizes human oversight, inclusiveness and transparency [16-18]. FDA lifecycle guidance similarly supports supervised software modifications and performance monitoring [19-22]. In resource-constrained settings, an interpretable rule using verified symptoms and structured imaging can be safer than a complex model with fragile external validity.

### **Governance and regulatory science**

Clinical research and clinical deployment are different legal and evidentiary activities. A prospective study can establish analytical and clinical validity, whereas deployment typically requires consideration of medical-device regulation, post-market monitoring, privacy and liability. Regulatory RWE policies illustrate the importance of traceability, prespecification, sufficient data quality and reproducible analyses [13-15]. The EMA's DARWIN EU network is a useful institutional example of how distributed observational analyses may support medicine regulation; analogous networks for endometriosis must still solve phenotype and imaging-data standardization.

### **Ethics and patient-centered benefit**

Early detection is only useful if it leads to better care. Any pathway that prioritizes a novel test while restricting access to experienced clinical assessment risks exacerbating inequality. Adolescents and patients who cannot undergo transvaginal imaging deserve

alternative appropriate options. Consent for reuse of intimate symptom histories, infertility records and imaging must be explicit and understandable. Subgroup monitoring should examine differential false-negative rates and delayed treatment among rural patients and those facing financial or geographic access barriers.

### **Limits of the present review**

This work is a structured narrative evidence synthesis, not an exhaustive systematic review. It does not have a registered protocol, complete citation-by-citation screening log, independent dual review, formal pooled reanalysis or access to confidential claims/EHR datasets. Database and journal names in methodological sections should not be mistaken for proof of comprehensive searches in all platforms. The selected evidence emphasizes highly relevant international guidance and landmark diagnostic studies; unavailable subgroup or post-deployment utility outcomes are treated as evidence gaps, not approximated as facts.

### **4.1 Uzbekistan: feasible digital diagnostic infrastructure**

#### **Clinical service and national policy**

For Uzbekistan, a clinically defensible strategy begins with gynecology and radiology workforce capacity, standardized referral criteria and quality assurance for pelvic imaging. The Ministry of Health has a coordinating role in standards, training, digital interoperability and safe data governance. However, no unpublished national prevalence, diagnostic-delay statistic or uptake figure is asserted here. Publicly available documentation would need separate verification before claims about nationwide coverage, electronic prescription completion or registry maturity are made.

#### **Electronic health and e-prescribing**

Electronic health records and e-prescribing initiatives could make endometriosis pathways visible if symptom onset, clinical suspicion, imaging phenotype, prescriptions and surgery are encoded consistently. The most informative near-term objective is not nationwide AI rollout but agreement on a minimum data set and interoperable report templates. A feasibility pilot should assess clinician workload, missing-data rates, patient consent and continuity of records across primary care and referral hospitals. Prescription data should retain distinctions between ordered, dispensed and reported medication use.

### **Registers, training and quality systems**

A phased endometriosis registry could start at selected tertiary and regional centers with structured IDEA-aligned imaging forms, standardized anatomy, treatment dates and outcome instruments. Governance should include a data protection assessment, access controls, version history and audit logs. Inter-observer reliability can be monitored using case review and second reading of a sample of scans. Rural facilities may initially need tele-education and referral networks rather than expensive stand-alone molecular testing systems.

### **AI implementation and evaluation**

Pilot AI could first perform low-risk functions such as checking completeness of referral records or flagging missing structured imaging fields. More consequential image-based triage or biomarker classifiers should require independent Uzbek-language and locally representative external validation, prespecified performance thresholds, calibration, subgroup assessment and prospective evaluation before informing care. Real-world monitoring must detect changes in scanner type, referral populations or coding practices. A regulator-approved research protocol should not be portrayed as approval for routine autonomous diagnosis.

### **Actionable implementation sequence**

Phase 1: multidisciplinary consensus on case definitions, reporting templates and clinical escalation rules. Phase 2: prospective multicenter observational registry with data-quality indicators and privacy safeguards. Phase 3: standardized imaging training and blinded inter-reader audit. Phase 4: independent evaluation of proposed biomarker or AI adjuncts against expert imaging pathways with patient-relevant endpoints. Phase 5: cost-effectiveness, equity impact, maintenance and post-deployment surveillance. This staged approach protects resources while enabling scientifically interpretable studies.

### **5. Conclusion**

Earlier and safer diagnosis of endometriosis depends on recognizing clinical suspicion before lesions become radiologically apparent, applying expert and appropriately chosen imaging, and avoiding the misuse of isolated biomarker results. Ultrasound and MRI provide established value for ovarian and deep disease but cannot reliably exclude

superficial peritoneal endometriosis; salivary miRNA signatures constitute an important, still context-sensitive validation advance. RWD and RWE can evaluate pathway delays, equitable access, longitudinal effectiveness and safety, whereas target-trial design, bias control and transparent AI validation are indispensable for credible causal and predictive claims. Uzbekistan can make substantive progress through standardized phenotyping, imaging training, interoperable and privacy-preserving registries, and independent prospective validation before adopting molecular or automated screening as routine practice.

### **Declarations**

Article type: Structured critical narrative review. Original patient data: None. No novel statistical estimates were computed. Funding: Not declared; to be completed by author(s). Conflicts of interest: Not declared; to be completed and signed by author(s). Ethics approval: Not applicable to this literature-based review. Author contributions and data availability: To be completed by the submitting author(s).

### **References**

1. World Health Organization. Endometriosis. Fact sheet. 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/endometriosis>
2. Becker CM, et al. ESHRE guideline: endometriosis. Hum Reprod Open. 2022;2022(2):hoac009. Available from: <https://academic.oup.com/hropen/article/2022/2/hoac009/6537540>
3. European Society of Human Reproduction and Embryology. Endometriosis guideline. 2022. Available from: <https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Endometriosis-guideline>
4. Validation of a Saliva Micro-RNA Signature for Endometriosis. NEJM Evid. 2025. doi:10.1056/EVIDoa2400195. Available from: <https://pubmed.ncbi.nlm.nih.gov/41147827/>
5. Transvaginal ultrasound and magnetic resonance imaging in detecting rectosigmoid deep infiltrating endometriosis: a comparative meta-analysis. 2025. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11955469/>
6. Guerriero S, et al. Systematic approach to sonographic evaluation of the pelvis in women with suspected endometriosis: IDEA consensus. Ultrasound Obstet Gynecol. 2016;48:318-332. Available from: <https://pubmed.ncbi.nlm.nih.gov/27349699/>

7. Guerriero S, et al. Addendum to IDEA consensus: sonographic evaluation of superficial endometriosis. *Ultrasound Obstet Gynecol.* 2025;66:541-547. Available from: <https://pubmed.ncbi.nlm.nih.gov/40632542/>
8. Guerriero S, et al. Addendum to IDEA consensus: sonographic evaluation of the parametrium. *Ultrasound Obstet Gynecol.* 2024;64:275-280. Available from: <https://pubmed.ncbi.nlm.nih.gov/38057967/>
9. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis and endometriosis classification systems: an international consensus statement. *J Minim Invasive Gynecol.* 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38819341/>
10. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis: an international consensus statement. *Ultrasound Obstet Gynecol.* 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38808587/>
11. Mick I, et al. Diagnostic Accuracy of Transvaginal Ultrasonography for Endometriosis According to the IDEA Consensus. *O G Open.* 2025;2:e061. Available from: <https://pubmed.ncbi.nlm.nih.gov/41000566/>
12. International Society for Gynecologic Endoscopy. Structured report for dynamic ultrasonography in suspected endometriosis. 2021. Available from: <https://pubmed.ncbi.nlm.nih.gov/34242934/>
13. FDA. Considerations for the Use of RWD and RWE to Support Regulatory Decision-Making. 2023. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug>
14. FDA. RWD: Assessing EHR and Medical Claims Data to Support Regulatory Decision-Making. 2024. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory>
15. FDA. Data Standards for Drug and Biological Product Submissions Containing RWD. 2023. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data>
16. WHO. Ethics and governance of artificial intelligence for health. 2021. Available from: <https://www.who.int/publications/i/item/9789240029200>

17. WHO. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. 2024. Available from: <https://www.who.int/publications/b/70584>
18. WHO. Artificial Intelligence for Health. 2024. Available from: <https://www.who.int/publications/m/item/artificial-intelligence-for-health>
19. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for AI-Enabled Device Software Functions. 2025. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence>
20. FDA. Artificial Intelligence-Enabled Medical Devices. 2026. Available from: <https://www.fda.gov/medical-devices/digital-health-center-excellence/artificial-intelligence-enabled-medical-devices>
21. FDA. Guidances with Digital Health Content. 2026. Available from: <https://www.fda.gov/medical-devices/digital-health-center-excellence/guidances-digital-health-content>
22. FDA. Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles>
23. WHO. Call for experts: new WHO guideline on endometriosis. 2026. Available from: <https://www.who.int/news-room/articles-detail/call-for-experts-for-the-new-who-guideline-on-endometriosis>
24. ESHRE. Endometriosis diagnosis recommendations and algorithm. 2022. Available from: [https://www.eshre.eu/-/media/sitecore-files/Guidelines/Endometriosis/ESHRE-GUIDELINE-ENDOMETRIOSIS-2022\\_2.pdf](https://www.eshre.eu/-/media/sitecore-files/Guidelines/Endometriosis/ESHRE-GUIDELINE-ENDOMETRIOSIS-2022_2.pdf)
25. ESHRE. Endometriosis recommendations and rationale. 2022. Available from: [https://www.eshre.eu/-/media/sitecore-files/Guidelines/Endometriosis/ESHRE-GUIDELINE-ENDOMETRIOSIS-2022\\_1.pdf](https://www.eshre.eu/-/media/sitecore-files/Guidelines/Endometriosis/ESHRE-GUIDELINE-ENDOMETRIOSIS-2022_1.pdf)
26. FDA. FDA Issues Comprehensive Draft Guidance for Developers of AI-Enabled Medical Devices. 2025. Available from: <https://www.fda.gov/news-events/press->

announcements/fda-issues-comprehensive-draft-guidance-developers-artificial-intelligence-enabled-medical-devices

27. Indrielle-Kelly T, et al. Application of IDEA consensus in preoperative ultrasound and MRI. *Ultrasound Obstet Gynecol.* 2020;56:115-116. Available from: <https://pubmed.ncbi.nlm.nih.gov/31876340/>

28. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis and classification systems: international consensus. *Eur J Radiol.* 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38879203/>

29. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis: international consensus. *Hum Reprod Open.* 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38807551/>

30. Non-invasive imaging techniques for diagnosis of pelvic deep endometriosis: international consensus. 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38812884/>

31. WHO. WHO releases AI ethics and governance guidance for large multi-modal models. 2024. Available from: <https://www.who.int/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models>

32. FDA. Medical Device Software Guidance Navigator. Available from: <https://www.fda.gov/medical-devices/regulatory-accelerator/medical-device-software-guidance-navigator>