

Reproductive, Metabolic, and Cardiovascular Consequences of Premature Ovarian Insufficiency: A Critical Integrative Review of Clinical Evidence, Real-World Data, and Precision Care

Xolboyeva Mashhura Yoqutjon qizi

Email – yoqutjonovnam@gmail.com

Termez University of Economics and Service

Faculty of Medicine

Abstract

Background: Premature ovarian insufficiency (POI), defined by impaired ovarian activity before 40 years of age, represents a multisystem diagnosis rather than an isolated fertility disorder. Its consequences involve hypoestrogenism, intermittent ovarian function, impaired reproductive autonomy, skeletal vulnerability, cardiometabolic risk, and substantial psychosocial burden. **Objective:** To critically integrate contemporary evidence regarding reproductive, metabolic, and cardiovascular sequelae of POI, and to assess how real-world evidence and computational tools may improve longitudinal care. **Methods:** This critical integrative review prioritizes the 2024 international POI guideline, peer-reviewed systematic reviews, observational cohorts, and official guidance from health regulators. Published effect estimates are attributed to original reports; no new patient data, original meta-analysis, or causal effect estimation was performed. **Results:** The updated international guideline identifies broader multisystem consequences and emphasizes hormone therapy, individualized fertility counseling, and ongoing bone and cardiovascular risk assessment. A 2023 meta-analysis of 16 studies reported an adjusted hazard ratio of 1.35 (95% confidence interval 1.06–1.63) for composite cardiovascular events and 1.42 (1.17–1.66) for coronary heart disease among women with POI, although residual confounding and heterogeneous case definitions constrain causal attribution. AMH is not a stand-alone diagnostic substitute for menstrual history and appropriate gonadotropin assessment. Clinical prediction should distinguish spontaneous ovarian activity from treatment-associated outcomes and account for selection, immortal-time, and confounding-by-indication biases. **Discussion:** Real-world data can characterize diagnostic delays, therapy persistence, screening uptake, and long-term events but cannot automatically reproduce randomization. Machine-learning models require external validation, calibration, fairness analysis, and clinician oversight. **Conclusions:** POI care should integrate reproductive counseling with sustained cardiometabolic and skeletal surveillance, appropriately individualized hormone therapy, psychosocial support, and pragmatic registries. Regulatory-grade evidence and ethically governed digital infrastructure offer a route to evaluating implementation rather than overstating treatment causality.

Keywords: premature ovarian insufficiency; ovarian failure; hypoestrogenism; infertility; reproductive aging; cardiometabolic risk; cardiovascular disease; bone

mineral density; hormone therapy; real-world evidence; causal inference; clinical registries; artificial intelligence.

1. Introduction

Scope and clinical meaning

Premature ovarian insufficiency refers to loss or substantial impairment of ovarian activity before age 40. Unlike an irreversible model of menopause, ovarian activity may be intermittent. This distinction affects counseling about contraception, fertility, and diagnostic uncertainty. The 2024 ESHRE–ASRM–IMS collaboration recognizes reproductive, skeletal, cardiovascular, neurological, sexual, and psychological dimensions and recommends diagnosis based on menstrual disturbance with appropriately elevated FSH [1,2]. The guideline reflects an updated prevalence estimate of 3.5%, although estimates depend on ascertainment, population selection, and definitions [1].

Biological heterogeneity

The diagnostic label encompasses genetic, autoimmune, iatrogenic, idiopathic, and other causes. Heterogeneity matters because age at onset, residual ovarian activity, comorbidity, previous gonadotoxic therapy, and access to hormone treatment can affect observed outcomes independently of estrogen exposure. A study that combines chemotherapy-induced ovarian insufficiency with spontaneous idiopathic cases may produce misleading average estimates. Relevant biological pathways include follicular depletion or dysfunction, altered hypothalamic–pituitary feedback, vascular endothelial signaling, adipose distribution, lipid metabolism, insulin action, and accelerated bone remodeling. These mechanistic possibilities are not identical to proven clinical effects.

Why long-term outcomes matter

The significance of POI extends beyond infertility or menstrual symptoms. Early loss of endogenous ovarian steroid exposure changes the length of time during which bone, cardiovascular, metabolic, and potentially neurological systems operate under a distinct endocrine milieu. Nevertheless, simple estrogen-deficiency narratives overlook shared determinants such as smoking, genetic susceptibility, chronic inflammation, cancer treatment, socio-economic status, physical activity, and hypertension. Responsible interpretation must balance biological plausibility against epidemiological uncertainty and avoid turning observational correlations into treatment mandates.

Real-world data and evidence

Real-world data (RWD) are routinely generated records of care and patient experience: electronic records, prescriptions, registries, laboratory systems, claims, wearables, and patient-generated observations. Real-world evidence (RWE) is clinical evidence about use, benefits, and risks generated through appropriately designed analyses of such data. The terms should not be used interchangeably. A database is not evidence until its measurements, population, time origin, causal question, and sources of bias have been

scrutinized. FDA guidance addresses data fitness for regulatory applications, while EMA supports distributed observational studies through DARWIN EU [6,7,8].

RCTs versus routine care

Randomized controlled trials estimate comparative effects under explicit eligibility, intervention, and follow-up conditions. They offer protection against measured and unmeasured baseline confounding but may have limited follow-up, selective inclusion, small subgroup sizes, and idealized adherence. RWE can describe persistence, outcomes in underrepresented patients, and events occurring years later; however, treatment choice is not randomized and outcome capture may be incomplete. These are complementary rather than competing evidence streams. In POI, the absence of long-term randomized treatment comparisons should not be disguised by presenting retrospective registry associations as experimental proof.

Efficacy, effectiveness, and safety

Efficacy denotes performance of an intervention in controlled study conditions, whereas effectiveness concerns outcomes in ordinary clinical settings. Hormone therapy effectiveness may depend on formulation, adherence, prior contraindications, cost, monitoring, and interruptions. Fertility-treatment efficacy measured per initiated cycle cannot simply be compared with population-level live birth probability because denominators differ. Safety signals in routine data may reflect prescribing patterns, previous illness, and differential surveillance. The distinction requires prespecified outcomes and explicit comparators.

Regulation and clinical AI

FDA RWE guidance, EMA real-world evidence programs, and WHO guidance on ethical use of AI each address important but different questions [6–10]. Regulators do not endorse every algorithm trained on routine data. AI can potentially assist chart abstraction, identify missed follow-up, or predict service needs, but plausible text generation is not independent clinical verification. Public-health algorithms should be audited for calibration drift and disparate error rates. In young women with POI, misinformation about infertility and inappropriate exclusion of reproductive possibilities could cause direct harm.

2. Materials and Methods

Review design and transparency

This manuscript is a critical integrative narrative review with methodological proposals; it is not a registered systematic review or a primary cohort study. The core evidence set was identified through targeted web verification of official guideline sources and indexed article records, emphasizing the 2024 POI guidance, the 2023 cardiovascular meta-analysis, WHO ethical guidance, FDA RWD/RWE policy, and EMA DARWIN EU documentation. PubMed, Scopus, Web of Science, Embase, Cochrane Library, and leading general medical journals are recommended source platforms for reproducible future searches; access and screening of every named

database were not independently performed for this manuscript. Accordingly, no PRISMA flow counts are presented.

Eligibility and interpretation

Priority was assigned to clinical guidelines, systematic reviews, meta-analyses, clearly defined longitudinal observational studies, and methodological references published in 2020–2026, alongside indispensable earlier methods papers. Extractable information includes populations, age thresholds, POI definition, treatment exposure, comparator, outcome, time horizon, confounding approach, effect measure, and uncertainty interval. Non-peer-reviewed hypotheses are labeled provisional. Repeated publications from overlapping populations must not be treated as independent replications; unavailable data are not imputed in this review.

Critical appraisal

Guideline recommendations were separated from measured empirical outcomes, and exploratory biomarker research was separated from validated clinical practice. The GRADE-informed guideline approach can inform confidence in care recommendations but cannot be transferred automatically to an unrelated observational effect estimate. Selection bias, residual confounding, reverse causation, informative censoring, immortal time, competing risks, differential loss to follow-up, and assay changes were considered conceptually. Formal study-level risk-of-bias scoring or reanalysis was not performed.

Propensity score matching

Propensity score matching (PSM) estimates the probability of treatment conditional on measured pretreatment characteristics and matches treated women to comparators with similar scores. In a POI hormone-therapy cohort, candidate variables include age, cause of POI, baseline cardiovascular risk, prior treatment, and healthcare utilization. Balance should be evaluated using standardized mean differences rather than outcome p values; overlap must be examined. Matching cannot adjust unmeasured determinants of prescription such as patient preference, contraindications absent from coding, or clinician judgment.

Inverse probability weighting

Inverse probability weighting (IPW) reweights observed patients to construct a pseudo-population in which measured treatment predictors are balanced. Stabilized weights and truncation may improve precision when probabilities are extreme. Time-varying exposure and confounding affected by previous treatment demand marginal structural methods rather than a single baseline weight. Positivity failure is especially plausible when clinicians almost never prescribe estrogen to patients with particular contraindications; no weighting strategy can create unsupported comparisons.

Cox regression and Kaplan–Meier

Cox proportional hazards regression describes the association between covariates and an event hazard under proportionality assumptions. Kaplan–Meier curves describe observed time-to-event experience assuming noninformative censoring; neither

method itself controls confounding. POI studies should distinguish age as time scale from years since diagnosis and report sensitivity to competing risks. Outcomes such as pregnancy, myocardial infarction, osteoporosis diagnosis, and therapy discontinuation require separate event definitions and censoring rules. Adjusted hazard ratios should never be described as absolute risks.

Target trial emulation and causal inference

Target trial emulation begins by specifying eligibility, treatment strategies, allocation time, follow-up, outcomes, causal contrast, and analysis plan before examining observational records. For hormone therapy after POI diagnosis, a clinically meaningful target trial might compare early initiation against a defined alternative in otherwise eligible women, with treatment grace periods and explicit adjustment for evolving health status. Cloning, censoring, and weighting may handle treatment switching, subject to assumptions. A directed acyclic graph identifies potential confounders and mediators; adjustment for downstream variables can introduce bias rather than remove it.

Machine learning and deep learning

Machine-learning prediction models might estimate likelihood of a missed cardiovascular screening visit or future low bone density. Candidate algorithms include penalized regression, random forests, gradient boosting, and neural networks. Deep learning may process longitudinal data or imaging but requires larger, representative data and rigorous external validation. Discrimination alone (AUROC) is insufficient: calibration, decision curves, subgroup performance, clinical utility, and postdeployment monitoring are essential. A high-dimensional model predicting treatment allocation is not automatically an unbiased estimate of treatment effect.

NLP, large language models, and survival analysis

Natural language processing (NLP) may extract POI onset, symptoms, or fertility counseling from free-text notes, but clinical negation, historical mentions, translations, and misclassification require chart review. Large language models (LLMs) may assist structured abstraction but are unsuitable as autonomous diagnostic adjudicators without validation. Survival analysis should specify left truncation, censoring, recurrent events, and competing risks. Flexible parametric survival models are alternatives when proportional hazards fail. Across designs, prespecified negative controls and sensitivity analyses for missing-not-at-random mechanisms strengthen credibility.

3. Results: Evidence Synthesis and Clinical Interpretation

3.1 Diagnostic phenotype and the reproductive trajectory

POI should be considered in women younger than 40 with menstrual irregularity, absent menstruation, or symptoms of estrogen deficiency. Current guidance favors menstrual disturbance for at least four months together with elevated FSH above 25 IU/L; a repeat concentration may be indicated when diagnostic uncertainty persists [1,2]. The clinical question is not whether every young patient with low estradiol has

irreversible ovarian exhaustion, but whether the presentation warrants systematic evaluation while excluding pregnancy and alternative endocrine explanations. A historical “two mandatory FSH values” rule from older guidance should not be presented as the unchanged 2024 recommendation.

3.2 Etiologic stratification and reproductive counseling

Etiologic assessment may include history of chemotherapy or pelvic surgery, family history, autoimmune conditions, and genetic evaluation when clinically indicated. Genetic findings can influence counseling for relatives and may reveal additional medical surveillance needs. However, broad genomic testing without consent, counseling, and follow-up capacity can generate uncertain variants that burden patients. The reproductive care plan must explain that intermittent ovarian function and occasional spontaneous conception are possible, while there is no reliably established intervention that restores ovarian function in all affected women [1].

3.3 Fertility preservation and treatment pathways

For individuals at risk of iatrogenic POI, fertility counseling ideally occurs before gonadotoxic exposure when clinical circumstances permit. Cryopreservation strategy depends on age, ovarian reserve, pubertal status, timing, indication, and local expertise. For established POI, donated oocytes are a recognized route to pregnancy, but access, legal frameworks, financial barriers, and personal values vary substantially. The outcome of interest may be live birth per patient, per retrieval, or per transfer; these denominators must not be interchanged. Fertility-related quality of life is not adequately captured by pregnancy rates alone.

3.4 Endocrine and metabolic pathways

Estradiol affects hepatic lipid handling, adipose distribution, insulin sensitivity, vascular tone, and bone remodeling, providing plausible pathways connecting POI to metabolic health. Causal pathways are nevertheless bidirectional: adiposity and chronic inflammation can influence endocrine markers and the probability of diagnosis. Clinical metabolic surveillance should focus on established risk determinants rather than assuming a laboratory signature unique to all women with POI. Heterogeneity by age at onset, etiology, treatment exposure, and baseline adiposity deserves direct analysis rather than broad categorization.

3.5 Cardiovascular morbidity

The 2023 systematic review and meta-analysis included 16 studies with 40,549 POI cases and 1,016,633 comparators. Reported pooled adjusted hazard ratios were 1.35 (95% CI 1.06–1.63) for composite cardiovascular events and 1.42 (1.17–1.66) for coronary heart disease, with limitations in comparability and other endpoints [3]. These results support cardiovascular risk recognition but do not prove that all excess risk is causally attributable to estrogen deprivation. Absolute event rates, follow-up, comparator menopause age, and residual confounding remain critical.

3.6 Cardiovascular risk assessment

Cardiovascular evaluation should consider blood pressure, smoking, physical activity, body composition, lipids, glycemic status, family history, and individual treatment exposures. Younger age creates a risk-communication challenge: near-term absolute atherosclerotic event probabilities can be small while lifetime consequences remain relevant. Conventional prediction calculators may not represent POI explicitly, and applying general population thresholds without clinical context can underrecognize risk. Conversely, a diagnosis of POI should not imply automatic medication escalation in the absence of conventional indications.

3.7 Bone mineral density and fracture prevention

Prolonged hypoestrogenism can accelerate bone turnover and compromise peak or maintained bone mass. Baseline bone-density assessment may be appropriate at diagnosis under international guidance, particularly in the context of additional skeletal risk factors [1]. Interpretation should use age-appropriate reference principles, monitor changes on consistent instruments when feasible, and avoid equating a single scan with future fracture inevitability. Weight-bearing activity, adequate nutrition, smoking cessation, and endocrine replacement where indicated belong to comprehensive prevention; fracture-event evidence in young women is thinner than evidence for bone-density differences.

3.8 Neurological, sexual, and psychosocial health

Cognitive complaints, sleep disruption, vasomotor symptoms, sexual discomfort, altered self-concept, and psychological distress may coexist. These outcomes can arise from hormonal changes and the shock of diagnosis, but are also modified by concurrent illness, treatment burden, relationships, and access to counseling. Screening should be symptom-led and respectful rather than presuming a universal neurocognitive disorder. In research, standard patient-reported outcome instruments and context-rich qualitative work can complement administrative endpoints.

3.9 Hormone therapy and clinical objectives

Hormone therapy is recommended for many women with POI without contraindications to mitigate consequences of early estrogen deficiency, generally until the usual age of menopause with periodic individualized reassessment [1]. This recommendation derives from integrated biological, observational, and clinical evidence rather than definitive RCT proof for every long-term cardiovascular endpoint. Patients with an intact uterus require appropriate progestogen protection alongside systemic estrogen when prescribed. Dose, route, comorbidities, contraindications, and patient preferences should guide selection; hormone therapy is not contraception.

3.10 Treatment persistence and gaps in implementation

Prescription issuance does not establish drug ingestion or long-term exposure. A valid RWE analysis should distinguish first fill, adequate coverage, refill intervals, dose modification, switching, contraindication-related discontinuation, and physician-documented clinical reasoning. Persistency estimates are sensitive to grace-period

choice and pharmacy access. Reasons for nonuse may include fertility misunderstanding, cancer-related fear, cost, supply instability, or limited provider experience. Improvement programs should be evaluated against patient-defined goals rather than only the number of electronic prescriptions.

3.11 Cardiometabolic disease prediction across specialties

Diabetes, cardiovascular diseases, oncology, infectious diseases, and neurological disorders are relevant mainly as comorbidity and transferability contexts for POI care rather than as interchangeable treatment outcomes. An endocrine prediction model trained among women with chemotherapy-associated POI could inappropriately learn cancer-surveillance patterns, while an infectious-disease hospitalization model could learn healthcare access rather than intrinsic risk. Cross-disease ML techniques can inform engineering practice, but claims about efficacy in diabetes or oncology do not demonstrate model validity for POI.

3.12 Real-world electronic health records

Electronic health records can capture menstrual history, POI codes, FSH and estradiol assays, blood pressure, obesity, smoking, fertility referrals, and treatment initiation. Their weakness is fragmentation: an infertility center may have a detailed endocrine evaluation while a primary-care file contains only a broad amenorrhea code. Validation requires chart review across settings and explicit confirmation that code date approximates the actual onset. Data quality, clinician coding behavior, and time-varying assay methods directly affect risk estimates.

3.13 Electronic prescribing and pharmacy data

Electronic prescriptions document intention to treat and may identify contraindicated co-medications, formulation changes, and refill patterns. Dispensing records improve exposure ascertainment, although inpatient medicines and out-of-pocket purchases may remain missing. The strongest design combines prescription orders, dispensations, encounter notes, and treatment stop dates. Hormone replacement, contraceptive pills, and gonadotropin stimulation drugs must not be grouped into one undifferentiated hormone-exposure category.

3.14 Laboratory databases and longitudinal assays

Laboratory systems can relate FSH, estradiol, AMH, thyroid markers, lipid profiles, and glucose measurements to clinical timelines. However, hormonal contraception or hormone therapy can distort diagnostic interpretation. AMH reflects ovarian follicular activity but should not replace diagnostic clinical criteria and FSH confirmation [1]. Missing measurements often indicate that testing was not ordered, not random omission; complete-case analysis may therefore select highly symptomatic or affluent subgroups and bias disease-severity estimates.

3.15 National registries and longitudinal linkage

A clinically designed POI registry could link baseline diagnosis to infertility treatment, pregnancy outcomes, fractures, cardiovascular events, and patient-reported outcomes. Unique pseudonymous identifiers would permit longitudinal follow-up while

preserving data minimization. Standard definitions are indispensable: POI diagnosed at a single fertility clinic differs from incident cases captured in a population registry. Linkage error, migration, changes in coding systems, and systematic exclusion of private clinics can distort apparent incidence and treatment access.

3.16 Claims, pharmacovigilance, and patient-generated data

Insurance claims provide coded service utilization but often omit menstrual details and actual symptom burden. Spontaneous pharmacovigilance reports may surface unexpected adverse reactions but generally lack reliable denominators, so reporting counts cannot establish incidence. Wearables capture heart rate, sleep, and activity at high frequency yet are vulnerable to selection bias, sensor changes, and missing wear time. Mobile applications can support symptom tracking and follow-up but must be evaluated for privacy, accuracy, and inequitable smartphone access.

3.17 Machine learning, deep learning, and reinforcement learning

Supervised models may help stratify the probability of missed screening, poor medication persistence, or future recorded cardiovascular events. Deep sequence models can capture longitudinal patterns but can also exaggerate associations from highly informative visit frequency. Reinforcement learning has been proposed for adaptive treatment policies in various medical fields, yet its application to POI remains investigational and would require safe evaluation before suggesting autonomous dosing decisions. Algorithms must never assume they have identified a causal best treatment solely from historical prescriptions.

3.18 NLP and large language models in POI care

NLP can extract chronology and explain otherwise inaccessible reproductive history from notes, including timing of chemotherapy or menstrual recovery. LLMs may draft patient education or summarize complex guideline text, but hallucination and failure to distinguish historical from current conditions remain clinically meaningful. Quality assurance requires human-adjudicated validation sets, multilingual testing, version control, and secure access. An AI-generated diagnosis should not become the reference standard used to train a second AI model.

3.19 Pharmacovigilance and adverse drug reactions

For women receiving estrogen–progestogen treatment, adverse-event analysis must separate formulation, route, dose, initiation age, treatment duration, and underlying contraindications. Drug safety signal detection using disproportionality methods is hypothesis-generating rather than comparative risk estimation. Case reports can flag uncommon events but cannot support a hazard ratio without a valid population denominator and confounder adjustment. Clinical surveillance should document benefits, tolerability, and medication changes rather than interpreting all symptoms as adverse reactions.

3.20 Pharmacogenomics and personalized care

Personalized POI care is grounded first in cause, age, fertility goals, contraindications, clinical risk, and preferences. Genetic testing can assist etiologic characterization

where appropriate, but no broadly validated pharmacogenomic panel currently guarantees optimal hormone therapy outcomes in routine POI care. Biomarker-rich precision medicine should be tested against simpler clinically grounded strategies. Unvalidated biomarker thresholds risk unnecessary expenditure, false reassurance, and unequal access to treatment.

3.21 Uzbekistan: clinical data infrastructure

For Uzbekistan, a practical priority is a multicenter, ethically governed POI cohort incorporating gynecology, reproductive endocrinology, family medicine, cardiology, and bone-health services. National Ministry of Health programs involving electronic health records, electronic prescriptions, and digital health could potentially support this infrastructure, but nationwide completeness, POI-specific registry coverage, and interoperability should not be assumed without current administrative verification. Establishing common data dictionaries, minimum privacy standards, clear consent pathways, and auditable clinical definitions should precede predictive AI deployment.

3.22 Uzbekistan: implementation proposals

A staged program could begin with clinic-level validation of POI coding against menstrual criteria and laboratory records, followed by structured capture of etiology, fertility goals, HT use, bone scans, BP, lipids, and follow-up. A second stage could support electronic referral reminders and quality indicators, including proportion of confirmed cases counseled about fertility and cardiovascular risk. A third stage could perform outcome-linked observational research after independent ethics approval and external data-quality audit. These are proposals, not claims about programs already operating.

Domain	Clinical outcome	Principal caveat	Evidence
Reproductive	Intermittent ovarian activity; infertility counseling	Do not equate POI with absolute sterility	[1,2]
Cardiovascular	Composite events: adjusted HR 1.35 (1.06–1.63)	Observational association; residual confounding	[3]
Coronary	Coronary disease: adjusted HR 1.42 (1.17–1.66)	Different ascertainment and follow-up	[3]
Skeletal	Bone density monitoring and prevention	Surrogate measures are not fractures	[1]
Hormone therapy	Mitigation of estrogen-	No RCT proof for all long-term endpoints	[1,2]

	deficiency consequences		
Digital evidence	Longitudinal adherence and outcomes	Coding, missingness, confounding	[6–10]

4. Discussion

The causal interpretation problem

A coherent biological narrative does not eliminate confounding. Patients diagnosed earliest may differ systematically from those diagnosed late in education, access, family history, and symptom severity. If an observational study shows reduced cardiovascular events with hormone therapy, clinicians may have selected women with fewer contraindications and better baseline health. Conversely, treatment may have been preferentially assigned to severe estrogen-deficiency cases, potentially obscuring benefit. Causal diagrams, new-user active-comparator designs, quantitative bias analyses, and transparent sensitivity checks are needed to distinguish these scenarios. Claiming that hormone therapy fully normalizes cardiovascular risk would exceed the underlying evidence.

Cross-study comparisons

The updated POI guideline synthesizes broader clinical considerations than a single cardiovascular meta-analysis. Consequently, recommendations regarding fertility counseling and hormone replacement should be interpreted within multidisciplinary decision-making, while hazard ratios from cardiovascular cohorts refer to precisely defined populations and endpoints. Apparent discrepancies may result from different age thresholds, the treatment status of controls, treatment era, survivor bias, or differential ascertainment of myocardial infarction. The most useful synthesis exposes these differences rather than averaging incompatible estimates.

The denominator problem in reproduction

Fertility research risks conflating three questions: whether follicular activity resumes, whether a pregnancy occurs, and whether a healthy live birth follows. Clinically, these are not interchangeable endpoints. Women referred for assisted reproductive technology may be selected on prior fertility attempts, resources, partnership circumstances, and preference, introducing severe selection bias into comparisons with untreated populations. Fertility decisions are also preference-sensitive; a strategy with high statistical pregnancy yield may be unacceptable to a woman whose priority is genetically related parenthood or whose circumstances preclude immediate conception.

Bone density as a surrogate

Bone mineral density is a useful intermediate indicator, but fracture risk is affected by falls, trauma, bone architecture, lifestyle, and age-related competing exposures. Short-term increases in density following hormone therapy do not in themselves confirm a

proportionate future reduction in fracture events. Research should therefore distinguish dual-energy X-ray absorptiometry outcomes from clinically adjudicated fractures and consider the effects of coexisting glucocorticoid use, eating disorders, smoking, and autoimmune disease. Standardization of repeat scanning matters as much as advanced predictive modelling.

Cardiovascular phenotype and prevention

The cardiovascular evidence supports integrating POI into a preventive clinical conversation rather than treating the diagnosis as an independent cardiology prescription. Lipid management, blood pressure control, smoking cessation, aerobic activity, nutrition, and diabetes screening continue to depend on individual clinical context. It is scientifically inappropriate to infer from one pooled hazard ratio that every patient experiences the same relative or absolute risk. Strong studies should report age-specific absolute risk differences and competing outcomes, not only adjusted relative estimates.

Confounding by indication and survivorship

Iatrogenic POI exposes important causal traps. Chemotherapy can affect vascular outcomes, weight, and mobility independently of ovarian suppression. Analyses limited to women who survived long enough to enter a menopause clinic may overrepresent comparatively healthy survivors. An appropriately designed target trial would define an index date close to diagnosis, account for cancer type and prior treatment, and minimize post-baseline exclusion. Separate analyses for genetic, autoimmune, and treatment-associated POI could improve clinical interpretability.

Data quality and missing information

A missing laboratory result is often not random; it may mean the patient had no access, was being treated empirically, or had care outside the digital network. Multiple imputation assumes conditions that may not hold if the absence of measurement depends on unrecorded disease severity. Plausible sensitivity analysis should include missingness indicators, pattern mixtures, and clinician review. Biochemical normalization under therapy must not be mistaken for endogenous ovarian recovery. EHR normalization requires careful mapping of date, dose, units, and disease definitions.

Explainable and ethical AI

Explainability is not synonymous with validity. Feature-attribution plots can identify which variables drive an algorithm, but cannot establish causation or prevent dataset shift. A model could identify high-risk patients simply because they have more clinical encounters, thereby penalizing patients who lack access to care. Ethical deployment requires prespecified intended use, calibration across ethnic, regional, and socio-economic groups, transparent oversight, privacy controls, and the ability for clinicians to challenge the recommendation. WHO AI guidance emphasizes governance and accountability rather than autonomous substitution of medical judgment [9,10].

Clinical decision support and human oversight

A useful clinical decision-support system might remind a clinician to review fertility counseling, bone health, or cardiovascular risks after confirmed POI diagnosis. This task is more readily auditable than an opaque model recommending medication based on a composite risk score. Implementation should compare the system against usual care for meaningful outcomes: completed follow-up, counseling quality, medication reconciliation, patient satisfaction, and unintended overtreatment. Alert fatigue, incorrect diagnostic coding, and inappropriate automatic disclosure of sensitive reproductive information require active mitigation.

Regulatory science and DARWIN EU

FDA and EMA frameworks recognize that regulatory-grade RWE depends on an adequate scientific question, relevant data provenance, analytical transparency, and reproducible methods [6–8]. DARWIN EU demonstrates a federated architecture that allows analysis of observational datasets without equating that infrastructure with a randomized experiment. A POI study seeking to inform treatment policy would require a protocol, validated phenotyping, clear comparators, complete outcome definition, and privacy-preserving governance. Merely labeling an analysis as “big data” or “AI-assisted” does not elevate its evidentiary status.

Generalizability to Uzbekistan

Uzbekistan is likely to face heterogeneous access to fertility specialists, laboratory testing, bone densitometry, and long-term specialist follow-up across care settings. The extent of each gap should be measured rather than assumed. A useful first step is a service-readiness audit across urban and regional facilities with documented referral pathways. Registry variables should capture language preference, travel distance, payment method, follow-up accessibility, and patient-reported priorities. These contextual measures are essential to avoid importing prediction thresholds from overseas health systems without calibration.

Strengths and limitations of this review

The synthesis joins endocrine mechanisms, guideline-based care, cardiovascular epidemiology, observational causal methods, and responsible digital implementation. Its principal limitation is that it is not a comprehensive protocol-driven systematic review of every indexed study. No new patient cohort was assembled and no pooled estimate was calculated; the numerical cardiovascular findings are those published by the cited meta-analysis. The methodological chapters describe appropriate future research designs, not analyses performed in this manuscript. Any journal presentation should retain this distinction, including explicit limitations and transparent editorial disclosures.

5. Research Agenda and Practice Recommendations

Standardize the case definition

A minimum data specification should identify onset age, four-month menstrual history, FSH testing and its timing, exogenous hormones, pregnancy exclusion, POI cause, and

referral date. The objective is not maximal data collection but adequate ascertainment of the target clinical state. Gold-standard chart adjudication samples should be used to calculate positive predictive value of electronic algorithms before deploying them at scale.

Emulate clinically meaningful decisions

Research should prioritize choices patients and clinicians actually face: hormone therapy initiation versus clinically reasonable alternatives; timely fertility counseling versus delayed referral; structured cardiovascular monitoring versus opportunistic monitoring. Each analysis should pre-register eligibility, time zero, treatment assignment window, co-interventions, and patient-relevant endpoints. Restrict comparisons to clinical overlap and document circumstances where causal estimation is infeasible.

Capture longitudinal experience

One-time hormone measurements describe only a narrow interval. Repeated measures of symptoms, therapy coverage, reproductive goals, blood pressure, body composition, lifestyle, and bone health can describe changing needs. Patients should be able to correct incorrect records and view how their sensitive fertility data are used. This is a necessary component of trustworthy RWE generation.

Evaluate real implementation

A pragmatic study might examine whether automatic reminders increase documented counseling or completed bone-density referrals. The primary outcome should be selected before implementation, with balancing measures for unnecessary referrals, clinician workload, privacy incidents, and inequitable outreach. A stepped or cluster-randomized implementation design may be preferable to purely observational comparisons when operationally feasible.

Protect algorithmic validity

Models should be developed on training data partitioned by patient and assessed with temporally and geographically external datasets. Calibration plots and net-benefit curves should accompany area-under-curve reporting. Fairness audits should quantify sensitivity and missed-care differences across clinically and socially relevant groups. Continuous governance is required after software updates, shifts in laboratory assays, or new treatment practices.

Integrate patient-centered outcomes

Outcome frameworks should include informed decision-making, mental well-being, symptoms, sexual health, and treatment acceptability in addition to cardiometabolic markers and pregnancy events. Fertility counseling should clarify uncertainty without eliminating hope or creating unrealistic guarantees. Communication must acknowledge that the scientific uncertainty surrounding long-term treatment effects does not justify withholding indicated guideline-based care.

6. Conclusion

Premature ovarian insufficiency is a multisystem reproductive-endocrine condition with important consequences for fertility, skeletal integrity, metabolic health, and cardiovascular risk. The most defensible clinical approach combines accurate diagnosis, individualized fertility and psychosocial counseling, appropriate hormone therapy, and long-term preventive assessment. Published cardiovascular associations establish an important signal but require careful distinction from causal treatment effects. RWD/RWE can strengthen care by identifying treatment gaps and longitudinal outcomes, provided disease definitions, confounding, censoring, missing data, and governance are addressed. AI is best used initially for narrow, auditable assistance rather than autonomous endocrine decision-making. For Uzbekistan, validated clinical registries and interoperable follow-up pathways are a more defensible starting point than premature deployment of unvalidated prediction algorithms.

References

1. Panay N, et al.; ESHRE, ASRM, CREWHIRL and IMS Guideline Group on POI. Evidence-based guideline: premature ovarian insufficiency. Hum Reprod Open. 2024;2024(4):hoae065. doi:10.1093/hropen/hoae065.
2. European Society of Human Reproduction and Embryology. Guideline on premature ovarian insufficiency [Internet]. ESHRE; 2024. Available from: <https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Premature-ovarian-insufficiency>
3. Cardiovascular Events Among Women with Premature Ovarian Insufficiency: A Systematic Review and Meta-Analysis. Rev Cardiovasc Med. 2023;24(7):193. doi:10.31083/j.rcm2407193. PMID:39077000.
4. European Society for Human Reproduction and Embryology Guideline Group on POI. ESHRE guideline: management of women with premature ovarian insufficiency. Hum Reprod. 2016;31(5):926-937. doi:10.1093/humrep/dew027.
5. ESHRE, ASRM, CREWHIRL and IMS Guideline Group on POI. Evidence-based guideline: premature ovarian insufficiency. Climacteric. 2024;27(6):510-520. doi:10.1080/13697137.2024.2423213.
6. US Food and Drug Administration. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products [Internet]. FDA; 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>
7. European Medicines Agency. Real-world evidence framework to support EU regulatory decision-making: third report [Internet]. EMA; 2025. Available from: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-3rd-report-experience-gained-regulator-led-studies-february-2024-february-2025_en.pdf

8. European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU) [Internet]. EMA. 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. Geneva: WHO; 2024. ISBN:9789240084759.
11. US Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices [Internet]. FDA; 2025. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices>
12. Page MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
13. von Elm E, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296.
14. Benchimol EI, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
15. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758-764.
16. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020.
17. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424.
18. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664.
19. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
20. Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34(2):187-220.
21. Collins GS, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
22. Wolff RF, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58.
23. World Health Organization. WHO releases AI ethics and governance guidance for large multi-modal models [Internet]. WHO; 2024. Available from:

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

24. Guideline Central. ESHRE premature ovarian insufficiency guideline summary [Internet]. 2024 [updated 2026]. Available from: <https://www.guidelinecentral.com/guideline/4543296/>

25. Panay N, et al. Evidence-based guideline: premature ovarian insufficiency [Internet]. PubMed; 2024. PMID:39647506. Available from: <https://pubmed.ncbi.nlm.nih.gov/39647506/>

26. European Society of Human Reproduction and Embryology. Evidence-based guideline: premature ovarian insufficiency [Internet]. PubMed Central; 2024. PMCID:PMC11631070. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11631070/>

27. Panay N. Premature ovarian insufficiency: new data and updated guidance [conference abstract]. Hum Reprod. 2024;39(Suppl 1):deae108.122.

28. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models [Internet]. WHO; 2024. Available from: <https://iris.who.int/bitstream/handle/10665/375579/9789240084759-eng.pdf>

29. European Medicines Agency. Real-world evidence framework: experience with regulator-led studies February 2024 to February 2025 [Internet]. EMA; 2025. Available from: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-3rd-report-experience-gained-regulator-led-studies-february-2024-february-2025_en.pdf

30. World Health Organization. Ethics and governance of artificial intelligence for health [Internet]. Geneva: WHO; 2021. Available from: <https://www.who.int/publications/i/item/9789240029200>

Author declarations

Review type: critical integrative narrative review, not original patient research. Data availability: no patient-level data generated or analyzed. Ethics approval: not applicable to literature-based analysis. Funding, conflicts of interest, and individual author contributions: to be completed only after author verification. AI assistance disclosure: to be completed according to target journal policies. References 24–30 include overlapping versions or records of official resources and must be consolidated/replaced with independent peer-reviewed studies prior to submission.