**WHEL** WOMEN'S HEALTH EVIDENCE LAB

METHODS · VERSION 0.1 · MAY 2026

# A multi-source drug repurposing signal database for under-researched women's hormonal conditions

Veronica Agudelo Columbia University

Correspondence vla2117@columbia.edu
Project rediscover-coral.vercel.app
Code github.com/veronicas-world/WHEL


STATUS OF THIS DOCUMENT

WHEL is an early-stage, independent project. It has not been peer-reviewed, and the method described here has not been externally validated. This is a working draft, written to make the approach transparent and to invite criticism while the project is still taking shape. The validation study in Section 4 has not been run yet.

If you want to point to this work, please treat it as a version 0.1 draft and link to the project site rather than citing it as a finished paper.

ABSTRACT

Drug repurposing means finding a new therapeutic use for a compound that is already approved. It is an appealing strategy in disease areas where conventional drug development has stalled, because the safety and pharmacokinetic groundwork already exists. Women's hormonal and reproductive conditions are one of those areas. Endometriosis, adenomyosis, polycystic ovary syndrome (PCOS), premenstrual dysphoric disorder (PMDD), vulvodynia, and the menopausal transition are all common and burdensome, yet all are thinly covered in the research literature relative to how many people they affect.

WHEL (Women's Health Evidence Lab) is an aggregator that collects drug repurposing hypotheses for these six conditions. It pulls structured and unstructured data from five sources: PubMed, ClinicalTrials.gov, the FDA Adverse Event Monitoring System (AEMS, formerly FAERS), the Open Targets Platform, and condition-specific Reddit communities. Each signal is scored on five evidence dimensions, namely replication, source quality, specificity, biological plausibility, and consistency of direction, by Claude Opus 4.6 working against a published rubric, and is then placed in one of four confidence tiers. The May 2026 snapshot holds 281 scored signals across the six conditions, drawn from 2,228 source citations.

WHEL is a hypothesis-generation tool, not a clinical recommendation engine, and it is free to use. This document describes how the database is built: the ingestion pipelines, the scoring framework, the validation work that is planned but not yet done, and the limitations and biases that come with the approach.

SECTION 1

# Background

## 1.1 The women's health research gap

The underrepresentation of women in biomedical research is well documented. The NIH Revitalization Act of 1993 was the first federal mandate that required women to be included in NIH-funded clinical research. That date matters, because it means the field has had only about three decades to correct an imbalance that ran for much longer (Mazure and Jones, 2015).

The conditions that affect women most severely have tended to attract low research funding even after you account for how common they are. Endometriosis affects up to one in ten women of reproductive age, and the average delay between first symptoms and diagnosis is still seven to ten years (Nnoaham et al., 2011). PMDD has a clear cyclical pattern and can be severely disabling, yet it is usually managed with SSRIs prescribed case by case rather than through a settled treatment pathway. Vulvodynia remains one of the most under-funded chronic pain conditions in the United States.

The result is a loop that feeds itself. Poorly characterized mechanisms make a condition harder to study. Being hard to study makes it harder to fund. Thin funding keeps the mechanism poorly characterized. WHEL does not break that loop, but it does try to make better use of the evidence that already exists inside it.

## 1.2 Why drug repurposing

Repurposing asks a narrower question than conventional drug discovery: does an already-approved compound have a therapeutic effect that nobody has formally recognized yet? Because the compound is already approved, a great deal of work is already done. Its safety pharmacology has been characterized, its pharmacokinetics are known, and in most cases it is already manufactured at scale. That makes the path to clinical investigation faster and cheaper than starting from a new molecule (Pushpakom et al., 2019).

The premise behind WHEL is that, for under-studied conditions, the useful signals often already exist. They are just scattered. A secondary endpoint buried in a trial report, a pattern in adverse event filings, a mechanistic link in a pathway database, a recurring observation in a patient forum: on their own these are easy to miss, and nobody has pulled them into a single condition-specific view. WHEL is an attempt to do that pulling-together.

## 1.3 What WHEL is and is not

WHEL is a signal aggregator. It does not generate new clinical evidence. What it does is surface evidence that already exists and ask whether it might be relevant to one of these six conditions. Its outputs are structured starting points for qualified researchers to investigate further. They are not treatment recommendations, and the language used across the site is written to keep that distinction clear.

SECTION 2

# Architecture

WHEL has three parts: a set of automated ingestion pipelines, a Postgres database hosted on Supabase, and a Next.js web frontend deployed on Vercel. The code is open-source under the MIT license, and the aggregated signal data is released under CC BY 4.0.

## 2.1 Conditions covered

The current snapshot covers six conditions:

Endometriosis. A chronic inflammatory disease in which tissue similar to the uterine lining grows outside the uterus.

Adenomyosis. Endometrial tissue invades the uterine muscle wall, producing painful and heavy periods.

Polycystic ovary syndrome (PCOS). A heterogeneous endocrine disorder with hyperandrogenism, irregular or absent ovulation, and metabolic features.

Premenstrual dysphoric disorder (PMDD). Severe luteal-phase mood and physical symptoms that cause real functional impairment.

Vulvodynia. Chronic vulvar pain with no identifiable cause, often neuropathic.

Perimenopause and menopause. The hormonal transition around the end of menstruation, with vasomotor, mood, sleep, and metabolic symptoms.

These six were chosen for three reasons: how common they are, how large the evidence gap is relative to that prevalence, and whether it was practical to write condition-specific search queries that did not bleed into adjacent conditions. Vulvodynia was included on purpose as a stress test, because its formal evidence base is very sparse and the point was to see how the framework behaved when there was little to work with.

## 2.2 Data pipelines

Five automated pipelines populate the database. A sixth, EudraVigilance, is built but is not yet contributing signals to the current snapshot.

> PubMed `DIRECT RESEARCH`
>
> Queries the NCBI E-utilities API with condition-specific Boolean search terms. Results are filtered by article type, publication date, and relevance, then sent in batches to Claude Opus 4.6. The model works from a system prompt that asks for a structured signal record: compound, signal type, evidence direction, a short summary, a mechanism hypothesis, and the list of PMIDs. The JSON output is validated and converted into parameterized SQL using `ON CONFLICT DO UPDATE`.

> ClinicalTrials.gov `DIRECT RESEARCH`
>
> Queries the ClinicalTrials.gov v2 REST API for trials targeting each condition. Trial phase, status, intervention type, and any posted adverse event tables are captured and stored. The same Claude classification step produces the structured signal record.

FDA AEMS CROSS-CONDITION SIGNALS

Queries the FDA Adverse Event Monitoring System (AEMS) through the OpenFDA REST API, using a two-pass strategy. The first pass is narrow: female patients reporting gynecological or condition-relevant reactions while taking the candidate drug. The second pass is a baseline: general female-patient reports for the same drug. A reaction has to appear in at least two AEMS reports to be surfaced, which filters out one-off noise, and reactions reported only once are dropped. Reaction-frequency summaries go to Claude for cross-condition classification. Every signal links back to the live OpenFDA query URL, so the underlying data can be checked directly.

Open Targets Platform PATHWAY INSIGHTS

Queries the Open Targets Platform GraphQL API using EFO and MONDO ontology identifiers for each condition. It retrieves drug candidate associations, mechanism-of-action data, target-disease association scores built from genetic evidence, known drug-target interactions, Reactome pathway analyses, and differential gene expression. Claude analyzes the output for pathway-level repurposing hypotheses. A signal that comes from this pipeline alone, with no human or pharmacovigilance corroboration, is classified as Exploratory and labeled that way.

Reddit COMMUNITY FORUM REPORTS

Queries condition-specific subreddits (r/Endo, r/endometriosis, r/PCOS, r/PMDD, r/Menopause, r/Perimenopause, r/adenomyosis, r/vulvodynia) through Reddit's public JSON search API, with no authentication. It runs eight treatment-focused queries per subreddit, including phrases like "what helped," "off label," "anyone tried," and "worked for me," and collects up to 25 top posts per query, deduplicated by post ID. Claude reads the posts and identifies treatments that at least two distinct users mention independently, with a focus on off-label or unexpected use. Individual post URLs are stored as source records, and each permalink is checked for a `/comments/` substring to confirm it points to a specific post rather than a whole subreddit. The pipeline is looking for consistent patterns across many posts, not for individual anecdotes.

EudraVigilance IN DEVELOPMENT

Built but not yet contributing signals. It queries the European Medicines Agency adverse event database (dap.ema.europa.eu) through the Oracle BI Analytics API, with substance codes resolved against the public adrreports.eu substance table. Female-patient reaction data is filtered and grouped by condition. A free registered EMA account is needed for session authentication.

## 2.3 Database schema

The signal database runs on PostgreSQL on Supabase with Row Level Security. There are five core tables: `conditions`, `compounds`, `repurposing_signals`, `sources`, and `cross_condition_patterns`. Repurposing signals connect compounds to conditions through foreign keys. The `repurposing_signals` table holds the five scoring dimensions (replication, source quality, specificity, plausibility, direction), a computed total evidence score, the confidence tier, the effect direction, and human-readable level labels. A unique constraint on the compound-and-condition pair prevents duplicate signals and lets pipeline reruns upsert idempotently.

The sources table links to signals and carries source-type metadata (pubmed, faers, clinical_trial, reddit, opentargets). That source type decides which evidence-arm tab a signal appears under on each condition page. Sources are deduplicated by URL before they are stored.

### 2.4 Frontend signal routing

Which tab a signal appears under on a condition page is decided at render time, from the source_type of the signal's associated sources.

| SOURCE TYPE OR SIGNAL TYPE | DESTINATION TAB |
|---|---|
| pubmed, clinical_trial | Direct Research |
| faers, opentargets (when AEMS-corroborated) | Cross-Condition Signals |
| pathway_signal, caution_signal | Pathway Insights |
| reddit | Community Forum Reports |

SECTION 3

## Scoring framework

### 3.1 Why a language model is the scoring layer

WHEL uses Claude Opus 4.6 (claude-opus-4-6) as the scoring and classification layer for every signal across all five active pipelines. This is the single most consequential methodological choice in the project, so I want to be plain about it.

The signals being scored are not the same kind of thing as one another. A peer-reviewed randomized controlled trial abstract, a registry trial protocol, an adverse event frequency summary, a pathway-database mechanism description, and a Reddit thread are structurally different inputs. Scoring all of them against one consistent framework calls for a layer that can read free text, apply domain judgment, and return a structured score. I tested smaller and faster models while building the pipelines. They produced flatter, less discriminating scores, and the weakness showed up most on biological plausibility and consistency of direction, which are the two dimensions that depend most on reading the source closely. I chose Claude Opus 4.6 for its performance on multi-criteria reasoning. That choice also carries real dependencies: model versioning, prompt sensitivity, and the risk of hallucination. Section 6 covers those.

### 3.2 The five-dimension rubric

Every signal is scored independently on five dimensions. Each dimension runs from 0 to 2, so the maximum total is 10. Scoring is done by Claude Opus 4.6 against the full source content, not just the metadata.

| DIMENSION | SCORE 0 | SCORE 1 | SCORE 2 |
|---|---|---|---|
| Replication | Single source only | Two independent sources | Three or more independent sources, same direction |
| Source quality | Forum or anecdotal | Observational, registry, or pharmacovigilance | Peer-reviewed human study or trial |
| Specificity | Vague outcome (“improved,” “felt better”) | Symptom-specific outcome (pelvic pain, cycle regularity, mood lability) | Clearly defined condition-specific clinical endpoint |
| Biological plausibility | Unclear or absent mechanism | Broad but plausible mechanism | Well-characterized drug-target-pathway-disease fit |
| Consistency of direction | Mixed or conflicting | Mostly consistent | Clearly consistent across all sources |

## 3.3 Confidence tier mapping

The total score maps to one of four confidence tiers, and that tier is what shows on the signal card.

| TIER | SCORE RANGE | DESCRIPTION |
|---|---|---|
| Exploratory | 0–3 | Single-source, mechanistic, or low-specificity signals; for hypothesis generation only. |
| Emerging | 4–6 | Early-stage evidence with some corroboration or mechanistic support. |
| Moderate | 7–8 | Replicated findings with solid mechanistic rationale. |
| Strong | 9–10 | Highly replicated, well-characterized signals with consistent direction across multiple evidence types. |

## 3.4 Category-specific minimum standards

On top of the numerical rubric, each evidence arm has a minimum bar a signal must clear before it is admitted.

> Direct Research
>
> The highest bar. A signal needs at least one peer-reviewed human study with a clearly identified population, drug, outcome, and effect direction. Mechanistic-only signals with no human data are excluded. The quality criteria reward replication and outcome relevance rather than citation count.

Cross-Condition Signals

Hypothesis-generating by nature. A signal has to appear in at least two independent evidence domains, for example literature plus AEMS, or AEMS plus community discussion, with the same direction of effect and a plausible shared biological mechanism. Three or more formal source mentions with consistent direction also qualify. Vague similarity between two conditions is not enough; there has to be a documented shared pathway.

Pathway Insights

Easy to overinterpret, so the requirements are specific. There must be a named mechanism (mast cell activation, prostaglandin signaling, androgen receptor modulation, not generic “inflammation”), at least one known drug-target link, and at least one disease-pathway link. A pathway-only signal with no human or pharmacovigilance corroboration is classified Exploratory and labeled that way.

Community Forum Reports

The strictest framing-level guardrails. There must be at least five distinct posts from unique users that use specific exposure-and-outcome language. The framework asks for specificity, a clear direction, and diversity across users. Reposts, promotional content, and low-content comments are excluded. Replication is graded 0 to 2 by post volume: 0 for five to seven posts, 1 for eight to fourteen, 2 for fifteen or more. A signal with fifteen or more qualifying mentions and a consistent direction can reach Moderate only when it is also triangulated with a formal source. The pipeline also records the time period of the discussion and whether the signal persists or reflects a temporary spike.

## 3.5 Cross-cutting reliability checks

Every signal, whatever arm it belongs to, also has to pass five reliability checks before it is admitted to the database.

1. Outcome specificity. “Improved” is not enough. A qualifying outcome is a condition-specific clinical endpoint.
2. Effect directionality. Each signal is classified as one of: improves, worsens, mixed, or unclear.
3. Replication. One source is interesting. Two or more independent sources start to look like a signal.
4. Confounding assessment. Known confounders are flagged: drugs with several indications, forum populations on more than one concurrent therapy, and AEMS data that may reflect reporting bias rather than real incidence.
5. Denominator awareness. AEMS and community data do not give true incidence rates. They generate signals, and a signal from either source needs corroboration before it can rise above Emerging.

### 3.6 Guiding principle

**Frequency is not truth.**

A rare signal that is specific, repeatedly observed, and from a credible source can carry more evidential weight than a high-volume but vague pattern. The scoring framework is built to reward specificity, reproducibility, and triangulation over raw volume.

SECTION 4

## Validation plan

The model-as-classifier approach needs to be checked against human judgment, and I want to be clear that the current snapshot does not include that check yet. A validation pass is planned before the project is submitted for academic review. It will work as follows.

1. Sample. A stratified random sample of about 75 signals, balanced across the four confidence tiers and the four evidence arms.
2. Independent rating. Two human raters, one with a clinical psychiatry and women's health background and one with a pharmacology background, will re-score each sampled signal against the published five-dimension rubric. They will be blind to the scores the model assigned.
3. Concordance metrics. The analysis will report Cohen's kappa for tier agreement (model versus human, and human versus human), percent agreement at the tier level, and the absolute deviation in summed score between the model and the human mean.
4. Failure analysis. Any signal where the model and a human rater disagree by two or more tiers will be reviewed by hand, to find systematic error modes, for instance the model leaning too hard on mechanism plausibility relative to replication.
5. Reporting. The full validation results, including the raw rating data, will be released alongside a methods update. If agreement comes in below an acceptable threshold (kappa under 0.6 at the tier level), the rubric and prompt structure will be revised before the next snapshot.

SECTION 5

## Related work

WHEL sits alongside several existing lines of work. A fuller Related Work writeup is on the project site.

#### PATIENT-GENERATED HEALTH DATA FOR ENDOMETRIOSIS

Phendo and Citizen Endo, from the Columbia University Department of Biomedical Informatics (McKillop, Mamykina, Elhadad, and colleagues), show that structured patient self-tracking can produce real scientific insight into endometriosis phenotypes and trajectories. WHEL's Community Forum Reports arm is conceptually close to this but methodologically different. Phendo collects structured, prospective self-

reports under ethical oversight. WHEL ingests retrospective, unstructured public discussion as hypothesis-generating signal. The two are complementary rather than competing.

#### DRUG REPURPOSING KNOWLEDGE BASES

DrugBank, the Open Targets Platform, RepoDB, and RepurposeDB provide drug-target and drug-indication associations at scale. WHEL ingests Open Targets directly (Section 2.2) and is structurally compatible with the others.

#### DRUG REPURPOSING METHODOLOGY

Pushpakom et al. (2019) and Hurle et al. (2013) set out the frameworks WHEL works within.

#### PHARMACOVIGILANCE SIGNAL DETECTION

Disproportionality methods for AEMS, such as the proportional reporting ratio, the reporting odds ratio, and the information component, are well established (Bate and Evans, 2009). WHEL does not reproduce them. The AEMS pipeline works from raw report-frequency summaries plus model-mediated cross-condition reasoning. Triangulating against disproportionality scores is on the development roadmap.

#### QUANTIFYING THE WOMEN'S HEALTH RESEARCH GAP

Mazure and Jones (2015) and the Society for Women's Health Research have documented the structural underrepresentation that motivates this project.

---

SECTION 6

## Limitations

A standalone Limitations section is published on the site and kept up to date. What follows summarizes it as of version 0.1.

#### METHODOLOGY

> **Model classification risk**
>
> All evidence scoring is done by Claude Opus 4.6. The risks of hallucination and prompt sensitivity are real and cannot be fully removed at the current state of the art. They are reduced by structured rubrics, JSON-schema-validated output, and the planned validation pass (Section 4). They are not eliminated.

> **Model versioning and prompt drift**
>
> Model versions and system prompts change over time. Each pipeline run pins a model version and the prompts are snapshotted in the code repository, but reproducibility across model upgrades is not guaranteed.

#### SOURCE-SPECIFIC BIAS

### Publication bias

PubMed indexes positive results and English-language journals disproportionately, so literature-sourced signals can be biased toward favorable findings.

### AEMS reporting bias

Spontaneous reports are subject to the Weber effect, channeling bias, indication confounding, and notoriety bias. AEMS signals are hypothesis-generating only. They are not causal estimates.

### Reddit community selection bias

Reddit users skew young, English-speaking, and treatment-frustrated. The Community Forum Reports arm captures what motivated communities talk about, which is not the same as what actually helps a representative population.

### Cross-condition signal interpretation

A drug that surfaces for a different indication in a woman who also has a target condition may reflect comorbidity rather than any pharmacological action on the target condition.

### Pathway inference is weak

Genetic-target overlap and pathway-level association are weak bridges to a real clinical effect. Pathway-only signals are classified Exploratory for that reason.

## SCOPE AND GENERALIZABILITY

### Sex-disaggregated data gap

Many PubMed and ClinicalTrials.gov entries do not report outcomes broken out by sex. Some signals labeled relevant to women are inferred from mixed-sex populations.

### Generalizability is not stratified

The current database does not stratify by race, age band, or geography. Signals reflect aggregate patterns that may not hold for specific subpopulations.

### Geographic and language scope

The current database is English-only. EudraVigilance integration (Section 2.2) will widen European coverage but not language coverage.

Scope limitation

Six conditions, not all of women's health.

OPERATIONAL AND DISCLOSURE

Temporal staleness

Pipelines run on demand, not continuously. Last refresh: May 2026.

Conflict of interest

WHEL receives no funding from the pharmaceutical industry. The author declares no commercial conflicts.

SECTION 7

## Data and code availability

Code. github.com/veronicas-world/WHEL, released under the MIT license. The repository holds the pipeline scripts, the Claude prompt templates, the database schema, and the Next.js frontend.

Data. A CSV and JSON export of the current signal database is planned but not yet posted. When it is, it will be released under CC BY 4.0.

Snapshots. Major snapshots are intended to be deposited to Zenodo with a citable DOI. None has been deposited yet.

Reproducibility. Pipeline scripts log the model version, a prompt hash, the query timestamp, and the source-record IDs for every signal they generate.

SECTION 8

## Ethics

WHEL ingests only publicly accessible data. The Reddit pipeline retrieves public forum posts through Reddit's public JSON API. No usernames are stored, and no post bodies are stored verbatim in the signal database; only post URLs are kept, as source records. No personally identifying information is collected or republished. The project does not meet the definition of human-subjects research under 45 CFR 46 and is not under institutional review board oversight. Even so, it follows the principles of transparency, minimum-necessary data use, and respect for the communities the data comes from.

SECTION 9

## Acknowledgements

WHEL is a solo project, but it has not been built in isolation. I am grateful for informal clinical feedback from my mother, a practicing psychiatrist, which has shaped how the conditions and their signals are described, and to the patient communities whose public discussion forms one of the project's evidence streams.

SECTION 10

## References

A fully formatted reference list will be assembled before any formal release. The working list is below.

Bate A, Evans SJW. Quantitative methods for pharmacovigilance signal detection. Pharmacoepidemiol Drug Saf. 2009;18(6): 427–436.

Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: from data to therapeutics. Clin Pharmacol Ther. 2013;93(4):335–341.

Mazure CM, Jones DP. Twenty years and still counting: including women as participants and studying sex and gender in biomedical research. BMC Womens Health. 2015;15:94.

McKillop M, Mamykina L, Elhadad N, et al. Designing in the dark: eliciting self-tracking dimensions for understanding enigmatic disease. Proc CHI. 2018.

Nnoaham KE, Hummelshoj L, Webster P, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril. 2011;96(2):366–373.

Ochoa D, Hercules A, Carmona M, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. Nucleic Acids Res. 2023;51(D1):D1353–D1359.

Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58.

NIH Revitalization Act of 1993, S. 1, 103rd Congress (1993).

Version 0.1 · May 2026 · Working draft · Comments welcome at vla2117@columbia.edu