

Supplementary Materials

Companion to: *Single- versus two-stage AI citation screening for clinical systematic reviews: a prevalence-aware, access-limited cost analysis* (Ivakhno, 2026).

All tables are derived from the released per-record data and the step CSVs; no value is hand-entered. Sections S1–S7 below; full rendered audit tables (S1, S3) and a search-strategy example (S2) follow in the Appendix.

S1 — Reference-include coverage and fate (audit trail for recall & β_{gold})

supplement/supp_S1_reference_include_coverage.csv — one row per published included trial across all four reviews (49 trials), with: corpus, trial label, PMID/DOI/NCT, whether it was **surfaced** by the search, its **AI screening decision**, and whether its **full text was retrieved**. This is the auditable basis for gate recall and β_{gold} .

Summary: 46/49 surfaced (3 misses all non-PubMed-indexed: Soliman [C], Faisal & Podder [D]); 43/46 surfaced trials screened in (strict); 35/46 full-text retrievable. Per-corpus recall and β_{gold} derive directly from these rows.

S2 — Verbatim search strategies

supplement/supp_S2_search_strategies.md — the exact Boolean queries persisted per corpus and per source (PubMed, ClinicalTrials.gov, Cochrane CENTRAL, Europe PMC), with date windows. Provided for search-reproducibility, as standard for systematic reviews. (Europe PMC returned 0 in these runs — a known term/API issue — so retrieval came from PubMed + CENTRAL + CT.gov + OpenAlex.)

S3 — Confidence calibration (supports H2)

supplement/supp_S3_confidence_calibration.csv — reliability table binning the gate's per-record confidence (deciles) against the observed gold rate per bin.

Headline: the model's mean confidence is ~ 0.98 while the overall gold prevalence is $\sim 4\%$; the expected calibration error (confidence treated as $P[\text{relevant}]$) is **ECE = 0.90**. The gold rate is near-flat across confidence bins — i.e. confidence does not track relevance — corroborating the H2 AUC result (raw confidence AUC 0.51). The score reports *decision-certainty*, not *probability of relevance*.

S4 — Cost-model sensitivity

step5_breakeven_grid.csv — the full (prevalence $\times$ cost-ratio) sweep of the cheaper strategy at $R^*=0.95$. Per-corpus summary (strict gate; cost in abstract-read units; $r_{\text{FT}}=1$):

Corpus	prevalence	pass- Φ	break-even c^*	single	two-stage	saving	feasible @R95
C vitamin-D	1.7%	0.17	1.20	2.50	1.42	43%	✗ (access)
B colchicine	3.1%	0.31	1.45	2.50	1.77	29%	✗ (access)
A ivermectin-Cochrane	3.6%	0.27	1.37	2.50	1.68	33%	✗ (access)

Corpus	prevalence	pass- Φ	break-even c^*	single	two-stage	saving	feasible @R95
D ivermectin-BMC	9.2%	0.31	1.45	2.50	1.78	29%	✗ (access)

The cost verdict (two-stage cheaper) holds for any cost ratio $> c^*$ (≈ 1.2 -1.45) and is therefore insensitive across the realistic range; "feasible @R95" is ✗ for every corpus because $\beta_{\text{gold}} < 0.95$ (the access ceiling), independent of the cost result.

S5 — Reproduced vs published study flow

Reconstruction used chiefly freely-queryable sources, so pools are smaller than the source reviews' multi-database screens (prevalence correspondingly higher). Drift is expected and reported, not hidden.

Corpus	Published: identified → de-dup → full-text → included	Reproduced: pool → AI-included → golds surfaced → full-text parsed
A ivermectin-Cochrane	570 → 382 → 182 → 11	311 → 84 → 11/11 → 248
B colchicine	1,152 → 989 → 21 → 5	161 → 49 → 5/5 → 157
C vitamin-D	5,733 → 2,483 → 83 → 8	416 → 70 → 7/8 → 374
D ivermectin-BMC	379 (+14) → 341 → 33 → 25	259 → 82 → 23/25 → 208

S6 — Gate failure-mode detail (supports H3)

The three strict-gate misses (all corpus D), with per-PICO-question judgements and confidence — Babalola (0.98), Biber (0.99), Niaee (0.74) — each marked population/intervention/outcome met but **comparator** unmet, hence a confident exclusion. Source rows are in `confidence_per_record.csv` (filter to these PMIDs/DOIs). Reported as a single observed pattern ($n=3$, one corpus, one PICO element), not a rate.

S7 — Reproducibility manifest

Released derived data (keyed by PMID/DOI; raw third-party text withheld):

- `confidence_per_record.csv` — per record: corpus, gold label, screening decision, confidence (0-1).
- `step3_gate_metrics.csv` — per corpus $\times$ gate: TP/FP/FN/TN, recall (Wilson CI), specificity, precision, pass- Φ .
- `step4_cost_inputs.csv` — β (OA availability) and c_a per corpus.
- `step5_corpus_decisions.csv` — per corpus $\times$ gate: prevalence, β_{gold} (per-trial), c^* , single/two-stage cost, feasibility at $R^* \in \{0.90, 0.95, 0.99\}$.
- `step5_breakeven_grid.csv` — (prevalence $\times$ cost-ratio) regime grid.
- `supplement/supp_S1_*.csv`, `supp_S2_*.md`, `supp_S3_*.csv` — as above.
- `make_figures.py` — regenerates all five main figures and computes the confidence AUCs from `confidence_per_record.csv`.

The screening model family/version and the eligibility/screening prompt configuration are documented in the release to support independent reproduction.

Not included here (require additional runs — see Limitations)

- Measured full-text *screening* recall (r_{FT}) and an $r_{FT} \in [0.7, 1.0]$ sensitivity sweep (no full-text-screening step was run; recall comparisons assume $r_{FT}=1$).
- A second screener model (all results are single-model).
- Full out-of-window-eligibility sensitivity across all four corpora (currently one corpus).
- Evaluation on an open benchmark (e.g. SYNERGY, CLEF-TAR, or the AgentSLR PERG dataset) for pipeline-independent comparison.

Appendix — rendered data tables

S1. Reference-include coverage and fate (all 49 published included trials). Trial labels truncated for layout; full labels are in the released CSV.

corpus	trial	PMID	DOI	NCT	surfaced	decision	full-text
A	Buonfrate 2021	34999239	10.1016/j.ijantimi cag.2021.106516		yes	included	yes
A	Chaccour 2021	33495752	10.1016/j.eclinm. 2020.100720	NCT0439002 2	yes	included	yes
A	Gonzalez 2021	35314650	10.3390/idr14020 020		yes	included	yes
A	Pott-Junior 2021	33723507	10.1016/j.toxrep.2 021.03.003		yes	included	no
A	Kirti 2021	34265236	10.18433/jpps321 05		yes	included	yes
A	Krolewiecki 2021	34189446	10.1016/j.eclinm. 2021.100959		yes	included	yes
A	Lopez-Medina 2021	33662102	10.1001/jama.202 1.3071	NCT0440584 3	yes	included	no
A	Mohan 2021	34483029	10.1016/j.jiac.202 1.08.021		yes	included	yes
A	I-TECH 2022 (Lim 2022)	35179551	10.1001/jamainter nmed.2022.0189	NCT0492094 2	yes	included	no
A	TOGETHER 2022 (Reis 2022)	35353979	10.1056/NEJMoa2 115869	NCT0472742 4	yes	included	yes
A	Vallejos 2021	34215210	10.1186/s12879-0 21-06348-5	NCT0452952 5	yes	included	yes
B	Lopes 2021	33542047	10.1136/rmdopen- 2020-001455	NCT0432256 5	yes	included	yes
B	Tardif 2021	34051877	10.1016/S2213-26 00(21)00222-8	NCT0432268 2	yes	included	yes
B	Deftereos 2020	32579195	10.1001/jamanet workopen.2020.1 3136	NCT0432679 0	yes	included	yes
B	RECOVERY Collaborative Group 2021	34672950	10.1016/S2213-26 00(21)00435-5	NCT0438193 6	yes	included	yes
B	Mareev 2021	33734043	10.18087/cardio.2 021.2.n1560	NCT0440324 3	yes	included	no
C	Entrenas Castillo M, et al. (Castillo), 2020	32871238	10.1016/j.jsbmb.2 020.105751		yes	included	yes
C	Elamir YM, et al., 2021	34508882	10.1016/j.bone.20 21.116175		yes	included	yes
C	Lakkireddy M, et al., 2021	34017029	10.1038/s41598-0 21-90189-4		yes	included	yes
C	Murai IH, et al., 2021	33595634	10.1001/jama.202 0.26848	NCT0444971 8	yes	included	no
C	Rastogi A, et al., 2020 (SHADE)	33184146	10.1136/postgrad medj-2020-13906 5	NCT0445924 7	yes	included	no
C	Sabico S, et al., 2021	34202578	10.3390/nu13072 170		yes	included	yes

corpus	trial	PMID	DOI	NCT	surfaced	decision	full-text
C	Sanchez-Zuno GA, et al., 2021	34071293	10.3390/jcm10112378		yes	included	yes
C	Soliman AR, et al., 2021		10.1177/20101058211041405		no	(not surfaced)	n/a
D	Ahmed 2021 — A five-day course of ivermectin...	33278625	10.1016/j.ijid.2020.11.191	NCT04523831	yes	included	yes
D	Babalola 2021 — Ivermectin shows clinical ben...	33599247	10.1093/qjmed/hcab035		yes	excluded	no
D	Biber 2021/2022 — The effect of early treatme...			NCT04429711	yes	excluded	yes
D	Buonfrate 2022 — High-dose ivermectin for ear...	34999239	10.1016/j.ijantimicag.2021.106516	NCT04438850	yes	included	yes
D	Chaccour 2021 — The effect of early treatment...	33495752	10.1016/j.eclinm.2020.100720	NCT04390022	yes	included	yes
D	Chachar 2020 — Effectiveness of ivermectin in...			NCT04739410	yes	included	yes
D	Beltran-Gonzalez 2022 — Efficacy and Safety o...	35314650	10.3390/ids14020020	NCT04391127	yes	included	yes
D	Krolewiecki 2021 — Antiviral effect of high-d...	34189446	10.1016/j.eclinm.2021.100959	NCT04381884	yes	included	yes
D	López-Medina 2021 — Effect of Ivermectin on T...	33662102	10.1001/jama.2021.3071	NCT04405843	yes	included	no
D	Mohan 2021 — Single-dose oral ivermectin in m...	34483029	10.1016/j.jiac.2021.08.021	CTRI/2020/06/026001	yes	included	yes
D	Ravikirti 2021 — Evaluation of Ivermectin as...	34265236	10.18433/jpps32105		yes	included	yes
D	Vallejos 2021 — Ivermectin to prevent hospita...	34215210	10.1186/s12879-021-06348-5	NCT04529525	yes	included	yes
D	Reis 2022 — Effect of Early Treatment with Iv...	35353979	10.1056/NEJMoa2115869	NCT04727424	yes	included	yes
D	Naggie 2022 — Ivermectin for Treatment of Mil...	36269852	10.1001/jama.2022.18590	NCT04885530	yes	included	no
D	Abd-Elsalam 2021 — Clinical Study Evaluating...	34076901	10.1002/jmv.27122	NCT04920942	yes	included	no
D	Bukhari 2021 — Efficacy of Ivermectin in COVI...		10.1101/2021.02.02.21250840	NCT04392713	yes	included	yes
D	Faisal 2021 — Potential use of azithromycin a...		10.29309/tpmj/2021.28.05.5867		no	(not surfaced)	n/a
D	Okumuş 2021 — Evaluation of the effectiveness...	33947344	10.1186/s12879-021-06104-9		yes	included	yes

corpus	trial	PMID	DOI	NCT	surfaced	decision	full-text
D	Podder 2021 — Outcome of ivermectin treated m...		10.3329/imcims.v14i2.52826		no	(not surfaced)	n/a
D	Shahbaznejad 2021 — Effects of Ivermectin in...	34052007	10.1016/j.clinthera.2021.04.007		yes	included	yes
D	Manomaipiboon 2022 — Efficacy and safety of i...	35982473	10.1186/s13063-022-06649-3		yes	included	yes
D	Lim 2022 — Efficacy of Ivermectin Treatment o...	35179551	10.1001/jamainternmed.2022.0189	NCT04920942	yes	included	no
D	Mancilla-Galindo 2022 (cited by Marcolino as...	36482326	10.1186/s12879-022-07890-6		yes	included	yes
D	Galan 2021 — Phase 2 randomized study on chlo...	33682640	10.1016/j.pathophys.2021.02.003		yes	included	yes
D	Niaee 2021 — Ivermectin as an adjunct treatme...		10.4103/1995-7645.318304	NCT04422561	yes	excluded	yes

S3. Confidence calibration (reliability table; ECE = 0.90).

confidence bin	n	mean confidence	observed gold rate
[0.0,0.1)	25	0.000	0.000
[0.1,0.2)	1	0.120	0.000
[0.2,0.3)	2	0.225	0.000
[0.3,0.4)	1	0.320	0.000
[0.4,0.5)	1	0.400	0.000
[0.5,0.6)	2	0.550	0.000
[0.6,0.7)	6	0.627	0.000
[0.7,0.8)	42	0.752	0.024
[0.8,0.9)	45	0.857	0.022
[0.9,1.0)	1022	0.979	0.043

S2. Search strategies (size-controlled).

The full verbatim per-source Boolean strategies (4 corpora × PubMed / CT.gov / CENTRAL / Europe PMC) are released as the data file `supp_S2_search_strategies.md`. To control PDF size, only the primary-source (PubMed) term for one corpus is reproduced here as an example; the remainder are in that file.

```
((("COVID-19"[MeSH Terms] OR "SARS-CoV-2"[MeSH Terms]) OR (COVID-19[tiab] OR COVID 19[tiab] OR SARS-CoV-2[tiab] OR 2019-nCoV[tiab] OR novel coronavirus[tiab])) AND (("Ivermectin"[MeSH Terms] OR (ivermectin[tiab] OR (ivermectin[tiab])) AND (randomized controlled trial[pt] OR randomized[tiab] OR randomly[tiab] OR placebo[tiab] OR controlled clinical trial[pt] OR clinical trial[pt])) AND ("2020/01/01"[dp] : "2022/12/31"[dp]) AND (randomized[tiab] OR randomised[tiab] OR randomly[tiab] OR placebo[tiab] OR double-blind[tiab] OR controlled trial[tiab] OR RCT[tiab]))
```