

Technical Supplement: Does New Medical Evidence Change AI Treatment Answers?

A Longitudinal Study of Evidence Citation, Medicine Naming and Selected Treatment Framing Around EASD 2026

Ozue C. Ezeife · AlphaCitation Research · October 2026 · Published by AlphaCitation Research, the research programme of AlphaCitation, a digital product of Akunudo LLC-FZ

1. Status and provenance

1.1 Identity of this document

Field	Value
Series	AlphaCitation Research Series, No. 1 (Paper No. 01)
Document	Technical Supplement
Version	v1.0.2 (v1.0: 12 October 2026; v1.0.1: 15 October 2026)
Post-publication notes	PN1, 15 October 2026: a change in the structure of Perplexity's responses between 6 and 7 October 2026 (paper §6.2; this supplement §4.1, §5.6 and §11.5.12(d)). Corrected 16 October 2026 (v1.0.2): §4.1 said that AlphaCitation's request did not change through the study, although the v5 setting changed it on 27 September (§5.1), and that every stored answer carried exactly 10 search results after the change; it now says that the request for scheduled study cycles and panel runs was the same on both sides of the 7 October change, that the bridges' v4-setting runs omitted the v5 setting by design, and that at least 98% of answers carried exactly 10 results on each day after the change. No count, verdict, method or figure changed
Date	October 2026
Author	Ozue C. Ezeife, Founder, AlphaCitation
Publisher	AlphaCitation Research
Peer review	None. This document has not undergone independent external peer review (§9.7). Human validation: blinded human validation of the interpretive coding (50 sampled E3, E5 and E3-i judgements), by the author as a single reader; not independent or external validation (§9.6).

1.2 The freeze

The single registered analysis ran once, on 11 Oct 2026 at about 16:57 UTC, at repository commit `d1fccaa9` (D20, with its time corrected by D21 item 4).

An audit of that frozen record, made while this supplement was written, found errata. Deviation D21 corrected each at its source. Every output was then regenerated on the same pull (67 cycles, sha256 unchanged) at commit `b2c52188` (2026-10-11 21:32:12 +0400). D21 states that no registered E2 verdict, registered E1 result, or E3, E5 or E3-i figure changed. Later commits that touched the frozen outputs:

- `1a502536`: per-cycle governed naming counts added to `easd-2026-analysis.json`; no figure changed.
- `66875044`: D23; the registered bridge self-check is computed and reported (`bridgeSelfCheck` in `easd-2026-analysis.json`); no registered verdict changed.
- `6d441b25`: the publication figures script.
- `818b54e6`: D24; post-registration robustness analyses, computed by a separate script (`robustness.ts`) into separate outputs. No registered verdict changed.
- `335baf2`, `84a2214c`: the publication figures script revised; no study output changed.
- `6392bc74`: the D24 outputs describe the local page only by neutral type (a finding of the second-round internal claim-to-evidence check, §9.7), and the per-answer table gains two columns (`citesLocalPageP61`, `echoesLocalPageSentence`); no figure changed.
- `7f0c974f`: the figures script, with a content gate on the page description.
- `89649d01`, `6d1b45d3`: the evidence-package builder and the release-wide scanner; no study output changed. `6d1b45d3` is the repository commit of the final freeze record.
- `a238941d`: the figures script (neutral cycle codes, Gemini gates); no study output changed.
- `672061ad`, `4659fe4c`: the seventh and eighth release-candidate builds of the evidence package, with the release tools revised after the freeze (the package builder and the release scan); no study output changed.

The record was then re-frozen. The figures in this supplement are copied from the regenerated files.

The freeze record `docs/evidence/computed/easd-2026-freeze.json` is written by `scripts/easd-2026-study/freeze.ts`. It records the role, path, size and sha256 of every input and output behind the published numbers, together with the sha256 of every study script.

The record copied into this supplement:

- `repositoryCommit`: `6d1b45d360c8d74cdd026cf7d2d7eb256e2d3e07`.
- `frozenAt`: `2026-10-12T09:50:57.567Z`.
- `problems`: `none`.
- Last committed in `1ef768ea` (2026-10-12 13:50:57 +0400).

Every hash in that record was recomputed against the working tree when this supplement was generated. 46 of 48 matched. The 2 that differ are release tools, kept outside the evidence package and revised after the freeze: `scripts/easd-2026-study/evidence-package.ts` (the package builder, revised after the freeze, last in `30123519`; frozen sha256 `b348d71c1eb622e0acc15472813a965be406f986a271b2878ab129bfff4c6df`, current sha256 `56324cdee10327a957f5c4b14b7261dc65d03b4ac195fd01242ed23a83f3deb4`, as `MANIFEST.json` (builder) records); `scripts/easd-2026-study/release-scan.ts` (the release scan, revised after the freeze, last in `4659fe4c`; frozen sha256 `8aeba6af218ea34bdaff0273de84f9c7400d239cee9d569ec8ee7c027f867153`, current sha256 `ae55ed219c75254a5c4d0e9557a9f8a805d527f0a072c331e7ab80824b84e3c5`, as the package README's version section records). They run no analysis, so the record was not re-frozen: every study script, input and output still matches. The package's `code/` holds only the scripts approved for release (§13.3).

The pull holds full question and answer text, so it is not in the repository. It is matched by its hash:

- File: `pull-final.json`, 110,898,974 bytes.
- sha256: `de98a604...eed265`.
- Pulled at `2026-10-11T16:56:27.066Z`.
- Covers 2026-08-22 to 2026-10-11: 67 cycles (`uae-glp1 23`, `singapore-sgl12 24`, `saudi-t1d 20`).

1.3 Registration and its timing

The study protocol was committed on 2 October 2026, following initial congress-period observations but before collection of post-congress observations.

The protocol docs/design/easd-2026-evidence-to-answer-prereg.md Was first committed as 8ee24e19 at 2026-10-02 13:50:05 +0400, which is **09:50:05 UTC on 2 Oct 2026**. Its own header says "~10:00 UTC"; D21 item 4 records the exact time. Sections 1 to 10 of the current file are byte-identical to that commit. Every later change is an appended entry in the §9 deviation log (git diff 8ee24e19 HEAD adds lines after line 205 only). One appended entry was edited after it was written: commit 1ee39872 changed "14:5x UTC" to "completed by 14:45 UTC" in the D14 addendum.

Placement in time:

Event	Time (UTC)	Relation to registration
Note 04 baseline published (the "before" named in the registered protocol)	24 Sep 2026	8 days before
Congress window opens (protocol §3)	28 Sep 06:00	4 days before
Last registered readout presented (ACHIEVE-4 LBA 25; proteomics LBA 29)	1 Oct 13:30	about 20 hours before
UAE during-window cycle UAE-C15 read before registration	completed 29 Sep 14:04	before (protocol header: "the last stored cycle read")
Registration commit 8ee24e19	2 Oct 09:50:05	—
Singapore during-window cycle SG-C17	completed 2 Oct 14:03	about 4 hours after
Congress window closes	2 Oct 22:00	about 12 hours after
First after-window cycle (UAE UAE-C16)	first capture 4 Oct 14:01; completed 4 Oct 14:05:03	about 2 days 4 hours after
First Saudi after-window cycle SA-C13	completed 4 Oct 16:34	after (the Saudi arm itself was added on 5 Oct, D10)

What the registration did and did not precede, stated plainly:

- It preceded every after-window cycle.
- It did not precede the congress. All seven registered readouts had been presented when it was written.
- It did not precede the UAE during-window cycle of 29 Sep. Those during-window naming counts are in the record the protocol rests on (docs/evidence/signal-note-06-post-easd.md), and were therefore known when the bands and rules were fixed.
- The registration timestamp is the git author and committer date of commit 8ee24e19 in AlphaCitation's repository (both identical). It is not a third-party registry timestamp.

Timing and interpretation. The protocol was committed before collection of after-window cycles, after the congress presentations and inspection of a during-window UAE cycle. Some computational interpretations were specified after preliminary after-window counts were visible; additional robustness and source-tracing analyses were undertaken after the results. These timings limit the confirmatory interpretation of the movement verdicts.

1.4 What is registered and what was added later

Element	Status	Where fixed
Question, sub-questions Q1-Q4, endpoints E1-E5, windows, admission rule, readout set (7 readouts), E2 decision and attribution rules, house magnitude scale, bridge design and negative control, reportable-result list, disclosure boundary	Registered	8ee24e19 §1-§10
Baseline coding scheme decisions S1-S11 (UAE, Singapore)	Fixed before any after-window cycle was coded; produced 2 Oct 10:08-10:14 UTC, committed 3 Oct (D0)	coding ledger §2
Readings I1-I5 of the analysis rules	Post-registration (D16), before the first analysis run	analyze.ts INTERPRETATIONS
OCTANE as a readout	Post-registration (D3)	D3
Saudi Arabia arm, its agent set, T1-T9	Post-registration (D10, D12, D13)	D10-D13
E3b (status errors about registered medicines)	Post-registration, exploratory (D6)	D6
Clustered intervals, strict scenarios, Saudi opportunity denominator, Saudi E3-i denominator, blind second pass	Post-registration robustness (D16)	D16
Full-precision E2 check, unverified-at-citation rule, freeze record	Post-registration reporting safeguards (D19)	D19
Bridge self-check (v4 same-day spread against band width)	Registered in §6; computed and reported only after the final run (D23); shown, not scored	D23
The three registered §5 descriptive source analyses	Registered in §5; reported late, under D24 (§11.5.1)	D24
Band-rule readings, cycle-level change, per-cycle view, corrected self-check, brand-only naming, search rates, attribution sensitivity, Saudi use-context recode, per-answer outcome table, E1 corrections, the local page's association	Post-registration robustness (D24), computed after the results were seen	D24
Analysis date of 11 Oct instead of 12 Oct	Post-registration, date only (D20)	D20
Errata after the freeze: Saudi E3 denominator, question text removed, OCTANE programme check and latency origin, corrected times and counts, notes, frozen verifier output, class (b) limitation	Post-registration correction (D21), after the final run; outputs regenerated on the same pull	D21

2. The registered protocol as committed, and every deviation

2.1 Registered text (commit 8ee24e19)

The passages below are quoted verbatim from git show 8ee24e19:docs/design/easd-2026-evidence-to-answer-prereg.md. Omissions inside a quotation are marked "...". The registered text does not name the measured medicines: §10 withheld them from the published note. This supplement names them generic-first — semaglutide (Wegovy), empagliflozin (Jardiance), teplizumab (Tzield) — under D22, the author's publication decision made after results were available (§2.3).

Objective (§1).

Did the evidence presented at EASD 2026 reach what four AI engines tell people about two established diabetes and obesity medicines in two markets — how fast, through which sources — and did it change how those medicines are named, positioned or described?

Question	Endpoint
Q1 Did congress evidence enter the answers, and how long after presentation?	E1 penetration + latency
Q2 Did the measured medicine's naming move beyond ordinary variation?	E2 naming movement
Q3 Did investigational agents presented at the congress enter answers to patients' treatment questions, and are they described as unapproved?	E3 pipeline intrusion
Q4 Did the clinical content of the answers change: the efficacy figures stated, and (Singapore) the place of SGLT2 inhibition relative to metformin?	E4 clinical currency · E5 first-line framing

Series, instrument and admission (§2).

Series	Market frame	Prompts	Answers / engine / cycle	Cadence
A GLP-1 medicine (UAE)	UAE, persona general	60	~54 scored	4 days (cron 14:00 UTC)
An SGLT2 inhibitor (Singapore)	Singapore, persona general	26	~19 scored	4 days

Engines and models, fixed: ChatGPT (gpt-4o-mini) · Claude (claude-haiku-4-5) · Gemini (gemini-2.5-flash) · Perplexity (sonar). **A cycle is admitted only if:** same four models; captures exist for ≥ 95% of its prompts; prompt set matches the series' first cycle at the rate already recorded (UAE 57–60/60, Singapore 26/26); zero cache hits. A cycle failing any of these is reported and excluded, never repaired.

The instrument changed on 27 Sep (ENGINE_VERSION 4 → 5). Sampling was pinned to a fixed setting on three of four engines; the fourth was already pinned. Models, questions and scoring did not change. ... No during- or after-window cycle is ever compared with a v4 figure except through §6.

Windows (§3, UTC).

Window	Definition	Expected cycles
Before, v4	Note 04's series through the last v4 cycle	UAE 13 (22 Aug – 25 Sep) · SG 15 (22 Aug – 26 Sep)
Before, v5	v5 cycles completing before 28 Sep 06:00	UAE 27 Sep · SG 27 Sep (excluded : 17/26 captured)
During	28 Sep 06:00 – 2 Oct 22:00	UAE 29 Sep · SG 2 Oct
After	cycles starting after 2 Oct 22:00 and completing by 11 Oct 23:59	UAE 4, 8 Oct · SG 6, 10 Oct (more if §7 cadence is approved)

The analysis is run once, on 12 Oct, over every admitted cycle in the windows above. No cycle after the cutoff is added to the published reading; a later cycle can only appear in a later note.

Readout set (§3), with programme times in CEST. The UTC equivalents used by the code are in §7.1.

REIMAGINE 1, LBA 11/12 (CagriSema) — Tue 29 Sep 14:45 · TRIUMPH-2 (retatrutide) — Wed 30 Sep 08:30 · CagriSema appetite, abstract 122 — Wed 30 Sep 14:30 · ACHIEVE programme (orforglipron) — Thu 1 Oct 08:30 · SMARTEST (SGLT2 inhibition vs metformin, first line) — Thu 1 Oct 09:30 · ACHIEVE-4 kidney outcomes, LBA 25 — Thu 1 Oct 15:30 · tirzepatide vs semaglutide proteomics, LBA 29 — Thu 1 Oct 15:30. **This list is the readout set.** A readout verified later is added under §9 and flagged as post-registration in every table that uses it.

E1, congress penetration and latency (§4; primary, both series).

Per engine, per cycle: the number of answers (of answers captured) that **cite** a congress-derived source, or whose **text reports a result** of a readout in the readout set.

Congress-derived source, four classes, decided by reading the fetched page, never the domain alone: (a) the congress's own pages, programme or abstracts; (b) a journal publication of a readout-set trial; (c) a sponsor release dated 28 Sep or later reporting congress data; (d) news or medical media dated 28 Sep or later reporting congress data. Every new URL in the window is classified and entered in a ledger with its fetch date and the sentence that qualifies it. Gemini redirect links are resolved by the `Location` header (read, not followed), as in the 29 Sep record.

Latency for each engine × readout: days from presentation to the first admitted cycle with ≥ 1 qualifying answer. No qualifying answer by the cutoff is reported as **not observed by day N** (right-censored), never as "the engine ignores the congress".

Beside the registered text: superseded interpretation. The registered "right-censored" reading of latency is superseded. This supplement reports a sampled detection bracket (last negative and first positive sampled cycles) and, where no qualifying answer was detected, not detected within follow-up (§7.1.4).

E2, naming movement (§4; primary, both series).

UAE: per engine, after-window pooled K/N of answers naming the medicine. Singapore: naming is at the ceiling (98.7–100% in every v4 cycle), so the E2 measure there is the governed **average rank** and the mention rates of the medicines named beside it.

Decision rule. A movement is declared for an engine only if its after-window pooled rate lies outside that engine's **bridged band**: the v4 single-cycle range (UAE: Perplexity post-step 46.3–58.8%, Gemini 48.1–74.1%, ChatGPT 20.4–44.4%, Claude 40.7–55.6%, four-engine 41.2–51.2%) shifted by the §6 v4→v5 offset for that engine, with each end widened by the offset's 95% interval. Magnitude is labelled on the house scale: < 5pp **SMALL** · 5 – < 10 **POTENTIALLY RELEVANT** · 10 – < 15 **DECISION-RELEVANT** · ≥ 15 **MAJOR**. If §6 does not run, E2 is reported **descriptively only**, and the note says so in those words.

Attribution rule, fixed now. A movement is called a *congress effect* only if, in the same engine and window, E1 is non-zero **and** the answers carrying congress-derived sources name the medicine at a rate different from that engine's other answers in the window. Every other combination is reported as what it is: *movement without congress sources, congress sources without movement, or neither*.

What the study can detect (95% half-width at p = 0.5): two after-window UAE cycles give ≈ 108 answers per engine, ± 9.4pp — only a **MAJOR** movement is detectable per engine. Seven daily cycles give ≈ 378, ± 5.0pp — a **DECISION-RELEVANT** movement becomes detectable. Singapore per-engine naming (≈ 19 per cycle) cannot resolve less than a **MAJOR** change on any plausible cycle count; that is why its E2 is positioning, not naming.

Correction beside the registered text. Singapore naming was near ceiling: the combined-engine cycle rate ranged from 98.7% to 100%, while individual-engine cycle rates ranged from 94.7% to 100% (15 v4 cycles, `perCycleNaming` in `easd-2026-analysis.json`).

E3, investigational-agent intrusion (§4; primary UAE, secondary Singapore).

Per engine, per cycle, across the **whole series from 22 Aug** (so the pre-congress rate is the comparator, on both instrument versions separately): answers naming an agent in the readout set that was investigational in the market frame on the day of the cycle — retatrutide, CagriSema (cagrilintide + semaglutide), and orforglipron in any market where it held no approval. Approval status per agent per market is established from the regulator's own record before classification, and recorded with its URL. Each naming answer is then read in full and coded:

- **Status correct** — states the agent is investigational, in trials, or not approved in the market;
- **Status absent** — names it without stating status;
- **Status wrong** — presents it as available, prescribable or approved where it is not.

The clinically material count is *status absent + status wrong in answers to a "which medicine" question*. It is reported however small, including zero.

E4, clinical currency of efficacy claims (§4; secondary, UAE).

Every answer that states a numeric weight-loss or glycaemic effect for the measured medicine is coded to its trial of origin. The reference values are taken from the medicine's label and the primary publications, fetched and recorded **before** any answer is coded, with URLs. Outcome per answer: *matches a labelled figure* · *matches a published trial figure not on the label* · *matches no identifiable source* · *matches a congress-window figure*. Reported per engine

per window. This is a description of what the answers tell patients, not a judgment of any engine.

E5, first-line framing (§4; secondary, Singapore; the SMARTEST readout).

Per engine, per cycle: answers describing SGLT2 inhibition (or the measured medicine) as a first-line option in type 2 diabetes, alternative to or ahead of metformin, coded by reading each candidate answer in full (candidate selection by text search for *first-line / first line / initial / instead of metformin*, then manual review). Before vs after, per engine. A rise with SMARTEST among the sources is reportable as such; a rise without it falls under the E2 attribution rule's wording.

Secondary, descriptive (§5).

The Dubai clinic page (Note 04's source): Perplexity citations per cycle, continued. New local or regional pages drawing $\geq 10\%$ of one engine's answers in a cycle ... Source-class mix per engine per window ... Domain churn is normal ..., so a new domain alone is never a finding.

The frozen analysis did not report these §5 analyses, and no deviation recorded the omission until D24. They are reported late, as registered descriptives, in §11.5.1.

Bridge (§6).

Purpose: convert the v4 band into v5 terms, so that Note 04's band can still judge the after window. **Design:** for each series, the stored battery (the exact prompt texts of the most recent admitted cycle), run on the same four models in **interleaved** replicates on one day: $R = 3$ at the v4 setting (each engine's vendor default, exactly as the v4 request paths sent it) and $R = 3$ at the v5 setting. Scored by the production scorer, unmodified.

Estimand: per engine, the v5 – v4 difference in naming rate, paired by prompt and pooled over replicates, with a bootstrap 95% interval over prompts. Also reported: v4-setting run-to-run spread ... and v5 run-to-run spread. ChatGPT's path did not change, so its offset is expected to be ≈ 0 ; **a ChatGPT offset whose interval excludes 0 means the bridge measured something other than the instrument, and the bridge is then reported as failed and not applied.** That is the bridge's built-in negative control.

The bridge is run once, between 3 and 9 Oct, and never re-run to obtain a different offset.

Cadence (§7). Both series move from 4-day to daily cycles from 3 to 11 Oct inclusive, subject to founder approval. The change is in cadence only.

Reportable results (§8).

congress evidence reached the answers in some engines and not others · it reached none of them by day N · naming moved without congress sources · congress sources arrived without moving naming · investigational agents appeared as options, with or without their status · answers' efficacy figures predate the congress, or match no source. **"Nothing moved, and nothing from the congress arrived" is a complete result and will be published as one.**

Disclosure boundary (§10).

Withheld: question text, battery composition, calibration figures, the parameter changed at v5 and its values, version history, regex-based diagnostics, scoring internals. Published: model names, denominators, windows, the readout set with its programme times, and the fact and effect of an instrument change, in neutral terms.

The amended disclosure boundary. The public disclosure boundary has been amended to release battery sizes and taxonomy, selected calibration results, version/deviation history, endpoint rules and approved coding diagnostics. Exact active question wording, the v5 parameter and its values, production scoring internals, unreleased implementation and full raw answers remain withheld. The amended boundary supersedes the conflicting descriptions in the earlier protocol; the original text remains available as provenance (§2.3).

- The publication names the three measured medicines (D22) and no clinic or local business. Withheld local entities appear only as typed tokens (§7.1.6, §13.3).
- The frozen diaTribe adjudication quoted one Saudi question verbatim until D21 item 2 removed it at source. The regenerated record describes the question without quoting it (§7.1.7).

2.2 Deviation log D0–D24

Commit times are the git dates of the commit that added each entry, converted to UTC. The *Observed* column uses four states:

- **pre-data:** no after-window cycle existed.
- **pre-coding:** after-window cycles existed, but no after-window answer had been coded or compared with a band.
- **counts seen:** after-window governed naming counts had been displayed.
- **prelim:** preliminary after-window results (coded tallies or naming counts) were visible.

D	Dated	Commit (UTC)	Observed	What changed	Why	Effect on any number
D0	3 Oct	7ca7e88f 3 Oct 09:05	pre-data	Records that the baseline coding (E3–E5, UAE 22 Aug–29 Sep, Singapore 22 Aug–26 Sep) ran 2 Oct 10:08–10:14 UTC, before the 2 Oct Singapore cycle existed; E4 reference values written before the coder first ran; cutoffs enforced in code	Timing record	None
D1	3 Oct	7ca7e88f	pre-data	E5 reading unit: every paragraph containing a candidate term plus the headings, not the whole answer	Practicality; applied identically to both windows	None on the comparison
D2	3 Oct	7ca7e88f	pre-data	E5 adds a <i>reported, not counted</i> category, "first-line with metformin" (48 of 202 baseline candidates)	It is the framing SMARTEST tests; collapsing it would hide movement	Protocol E5 count unchanged; category reported beside it
D3	3 Oct	7ca7e88f	pre-data	OCTANE added to the readout set, post-registration, as an E4 reference value and E1 class (c) source	A congress-window efficacy figure for semaglutide (Wegovy), UAE	OCTANE latency rows exist only because of D3; labelled post-registration
D4	3 Oct	7ca7e88f	pre-data	Every E4 figure scripted "no identifiable source" or "congress-window", both windows, is read by hand before publication	A 50-figure audit found 2–4 errors (92–96%), below the 95% bar	E4 no-source and congress columns are hand-checked (§7.6)
D5	3 Oct	7ca7e88f	pre-data	An agent listed as an option with its status stated stays <i>status correct</i> ; options counted descriptively	Coding clarification	None on E3 counts
D6	3 Oct	7ca7e88f	pre-data	E3b, exploratory: status errors about <i>registered</i> medicines, screened in both windows	The baseline showed an error class E3 cannot see	No E3b figure appears in the frozen analysis outputs

D	Dated	Commit (UTC)	Observed	What changed	Why	Effect on any number
D7	5 Oct	91a0634a 5 Oct 07:38	pre-coding (after-window cycles existed; none coded)	Two coder regular expressions carried a raw backspace byte (0x08) for \b; fixed, baseline regenerated from the same pull	Repository control-byte gate	Baseline E4: answers with a coded figure 452 → 450; "no identifiable source" answers 100 → 98; figures excluded by rule 96 → 97. E3, E5 unchanged
D8	5 Oct	91a0634a	pre-coding	Corrects D6's illustration (the "awaiting approval" sentences concern a tablet registered 4 May 2026)	Hand read	None
D9	5 Oct	e7f1536e 5 Oct 11:58	pre-coding	Daily cadence from 5 Oct (not 3 Oct), 5–11 Oct, both series; scheduler due-check fixed	\$7 approved by the founder on 5 Oct; deploy timing	More after-window cycles; E1 latency resolves to the day from 5 Oct
D10	5 Oct	0f979b76 5 Oct 12:36	pre-coding (Saudi 4 Oct cycle on file, unread)	Third arm added: teplizumab (Tzield), Saudi Arabia, 56 questions, same four models; E1 and E2 (own band, own bridge); E3 with a type 1 diabetes agent set; E4/E5 not applicable; labelled post-registration everywhere	Founder's request (Gulf materiality)	Every Saudi figure is post-registration and never pooled
D11	5 Oct	74c46c49 12:52, 58bf3664 12:52	counts seen (bridge validation displayed the UAE 4 Oct and Saudi 4 Oct per-model naming counts)	First UAE bridge attempt lost to a tooling kill after 4 of 6 runs (about \$24, nothing used); bridge rerun with a per-outcome checkpoint; design unchanged	Durability	None on design; cost reported (\$5.5)
D12	5 Oct	1818c97d 13:29	pre-coding for Saudi answers	Saudi E3 agent set, rules, regulator snapshot committed before any Saudi answer coded; Saudi status "not established", every code provisional; E3-i defined, cap 60 per window per agent, seed 20261005	D10 required it	Defines Saudi E3 and E3-i
D13	5 Oct	eac37257 13:33	pre-coding for Saudi answers	Saudi prompt-match rule corrected from 54/56 to 52/56 (the rate already recorded); persona label corrected to match the stored questions (the framing instruction itself is withheld; §4.2)	D12 had transcribed \$2 wrongly; it rejected 10 of 11 pre-congress cycles	Admits the Saudi cycles; Saudi 27 Sep v5 cycle excluded (9 of 56 captured)
D14	5 Oct	d3a2ee85 14:37; addendum 2368cdd1 14:45, 1ee39872 14:45	pre-coding (capture counts and timestamps only)	UAE 5 Oct cycle UAE-C17 excluded: 45 of 60 questions per engine served from a 24-hour cache; cache TTL cut to 12 hours; key namespace moved	\$2 admission rule	UAE after window has 7 admitted cycles (4, 6–11 Oct), one fewer than planned
D15	8 Oct	7ed02a5e 8 Oct 14:57	counts seen (bridge validation displayed Singapore and Saudi 7 Oct counts)	Singapore and Saudi bridges ran 8 Oct, not 6 Oct; D11's premise about the shared spend cap corrected; EDE directory outage recorded and a control check added to the regulator snapshot	No operator session 6–7 Oct; outage found in rehearsal	Bridge source cycles are 7 Oct; UAE regulator status carried forward from 2 Oct
D16	10 Oct	899b967f 10 Oct 16:27	prelim (hand-coded E3, E5 and Saudi E3-i tallies through 10 Oct; UAE per-engine naming counts through 10 Oct; no rate compared with a band)	Readings 11–15; E1 scope rule; post-registration robustness (clustered intervals, strict scenarios, Saudi opportunity denominator, Saudi E3-i denominator, blind second pass)	Registered words left computations open	Fixes how E2 bands, magnitude and attribution are computed; adds robustness columns
D17	10 Oct	4aaaa962 16:50; result eb349f87 16:55	prelim	D4 hand read excluded 104 of 195 figures; a seeded 150-figure audit of the 1,082 label/published figures is fixed with a 5% bar; result 8.0% (Wilson 95% 4.6–13.5%)	Scripted attribution far less reliable than D4's audit implied	E4 label/published split not published (quarantine)
D18	11 Oct	2c78560e 11 Oct 15:26	prelim	E1 sources' readouts assigned by reading the whole article; the diaTribe round-up re-labelled; readout data reaching another arm reported apart	An external reviewer's check found an unnamed TRIUMPH-2 item	No registered cell changes; cross-arm row added (§7.1.7)
D19	11 Oct	43f6c990 11 Oct 16:51	prelim	Full-precision E2 check; "unverified at citation" rule; freeze record	Reporting safeguards before the final run	Adds the full-precision table; TRIUMPH-2 in the diaTribe page is never counted
D20	11 Oct	d1fc9aa9 11 Oct 16:57	final run	Analysis ran on 11 Oct, not 12 Oct; UAE regulator status carried forward from 2 Oct. D20's text gives the run time as "about 17:15 UTC"; D21 item 4 corrects it to about 16:57 UTC (pull stamped 16:56:27Z, commit 16:57:47 UTC)	The window's last scheduled cycle had completed; a fresh pull matched cycle for cycle	Date only; no cycle added or removed
D21	11 Oct	b2c52188 11 Oct 17:32	after the final run (errata)	Errata found by the technical-supplement audit, each corrected at source, outputs regenerated on the same pull and re-frozen: (1) Saudi E3 denominator and opportunity count taken from the searched set, governed count shown for reference; (2) the quoted Saudi question removed from the diaTribe adjudication; (3) OCTANE verified against the Final Programme (LBA 59, LBA Short Oral Event C, Wed 30 Sep 11:45–12:45 CEST), latency origin moved from midnight to 09:45 UTC; (4) times corrected (D20 run about 16:57 UTC; registration 09:50:05 UTC; manifest built from the 16:36 pull); (5) counts stated outside the outputs corrected (D4 113 of 207; coverage range; Saudi bridge cost); (6) notes (A10 precedent dated 8 Oct; S7 stale, D3 governs; the cache cycle failed the capture rule; both second-pass batches seed 20261011); (7) verifier output frozen; (8) class (b) has no date condition, reported as a design limitation	Defects in the frozen record	Saudi E3 denominators 2,446 / 1,789 (were 2,402 / 1,757); OCTANE latency Perplexity 4.6 → 4.2, Gemini 6.6 → 6.2, censoring 11.6 → 11.2. No registered E2 verdict, registered E1 result, E3, E5 or E3-i figure changes
D22	11 Oct	db0f8fb1 11 Oct 17:45	after preliminary and final results	The flagship publication names the three measured medicines — semaglutide (Wegovy), UAE; empagliflozin (Jardiance), Singapore; teplizumab (Tzield), Saudi Arabia; clinics, local businesses and every other \$10 item stay withheld; \$10 and D22 reproduced in this supplement (§2.3)	Founder decision: a publication for scientific scrutiny, reproducibility, verification of regulatory context and practitioner use must say what was measured; not part of the registered protocol	None: no endpoint, numerical result, eligibility rule, statistical method, attribution criterion or registered interpretation changes
D23	11 Oct	66875044 11 Oct 18:13	after the final run	The \$6 bridge self-check (v4-setting same-day spread against the v4 band width, v5 spread beside it), omitted from the analysis, is computed from bridge.json and reported (bridgeSelfCheck); shown, not scored	An internal audit identified the omission of a registered check	None on any registered verdict. The check does not hold as worded: the same-day v4 spread is 0.101–0.497 of the band width in every UAE and Saudi engine (§5.6)

D	Dated	Commit (UTC)	Observed	What changed	Why	Effect on any number
D24	12 Oct	818b54e6 11 Oct 20:01 UTC (2026-10-12 00:01 +04)	after the final run and the internal quality gate	Post-registration robustness analyses computed by <code>robustness.ts</code> into separate outputs (band-rule readings, cycle-level change, per-cycle view, corrected self-check, brand-only naming, search rate, attribution sensitivity and minimum p, Saudi E3-i use context, per-answer outcome table, E1 corrections); the three registered \$5 descriptive analyses reported late; the local Gemini redirect cache added to the freeze record	An internal claim-to-evidence check and three internal perspective reviews (methodological, pharmaceutical, GEO) of 11 Oct found gaps; not published (§9.7)	None on any registered verdict or figure; every output labelled post-registration robustness (§11.5)

Mapping of commits to entries: D0–D6 in `7ca7e88f`, D7–D8 in `91a0634a`, D9 `e7f1536e`, D10 `0f979b76`, D11 `74c46c49` (with its disclosure paragraph in `58bf3664`), D12 `1818c97d`, D13 `eac37257`, D14 `d3a2ee85` (addendum `2368cdd1`, edited by `1ee39872`), D15 `7ed02a5e`, D16 `899b967f`, D17 `4aaaa962` (result `eb349f87`), D18 `2c78560e`, D19 `43f6c990`, D20 `d1fccaa9`, D21 `b2c52188`, D22 `db0f8fb1`, D23 `66875044`, D24 `818b54e6`.

2.3 Disclosure decision history: §10 as registered, and D22

The original disclosure boundary and its dated amendment are reproduced below. Both are copied by script from `docs/design/easd-2026-evidence-to-answer-prereg.md`: §10 from commit `8ee24e19`, and D22 from the current file.

§10, as registered (verbatim):

10. Disclosure boundary (unchanged from Notes 04 and 06 drafts)

Neither medicine, the clinic, nor any local business is named in the published note. Withheld: question text, battery composition, calibration figures, the parameter changed at v5 and its values, version history, regex-based diagnostics, scoring internals. Published: model names, denominators, windows, the readout set with its programme times, and the fact and effect of an instrument change, in neutral terms.

D22, as appended at `db0f8fb1` (verbatim):

D22 — 2026-10-11, the three measured medicines are named in the publication; §10 otherwise stands. §10 withheld the measured medicines from the published note. The founder decided on 11 Oct that AlphaCitation Research Series No. 1, the flagship publication that replaces the short note, names all three — Wegovy (UAE), Jardiance (Singapore) and Tzield (Saudi Arabia) — because a publication intended for scientific scrutiny, reproducibility, verification of regulatory context and broad practitioner use must say what was measured. **This editorial decision was made after preliminary and final results were available.** It is not pre-registered, and it is not neutral in perception: naming a medicine can be read as commentary on it, so the publication states that naming implies no endorsement, sponsorship or commercially meaningful effect, and discloses AlphaCitation’s interests and affiliations. The three are named consistently, whatever their results; none is emphasised for having a more interesting finding. Clinics and local businesses remain withheld, as do question text, battery composition, calibration figures, the v5 parameter and its values, version history, regex diagnostics and scoring internals. No endpoint, numerical result, eligibility rule, statistical method, attribution criterion or registered interpretation changes. §10 and this deviation are reproduced in the technical supplement so the full decision history is visible.

3. Registered versus exploratory: the endpoint map

Endpoint, analysis or readout	Status	Notes
E1 per-engine, per-cycle count of qualifying answers (source route + text route), UAE and Singapore	Registered	Classes (a)–(d), §4
E1 latency for the 7 registered readouts, own series	Registered	Sampled detection bracket; not detected within follow-up where no qualifying answer was detected
E1 latency for OCTANE	Post-registration (D3)	Exploratory; reported separately
E1 latency for "Other EASD 2026 data" (classes c/d outside the readout set)	Post-registration (label added in the E1 tally)	Qualification of such answers <i>is</i> registered E1. Only the per-readout latency row is post-registration
E1 cross-arm readouts (a readout reaching another series)	Post-registration (D18)	Never enters the registered latency table
E1 "unverified at citation" status	Post-registration (D19)	Never counted, never treated as negative
E1 inspection coverage, unreadable-domain counts	Post-registration reporting (E1 scope, D16)	Descriptive
E2 UAE per-engine and four-engine movement against bridged band	Registered (rule); computation readings I1–I3 (D16)	Verdicts as registered
E2 magnitude label	Registered scale; reading I2 (D16)	Descriptive label
E2 attribution (Fisher)	Registered rule; reading I3 (D16)	—
E2 Singapore average rank and competitor mention rates	Registered as descriptive (§4); reading I4 (D16)	No band, no verdict
E2 Saudi	Post-registration (D10, I5)	Exploratory
E2 full-precision distances, exact-cycle band	Robustness (D19)	Never reverses a verdict
E2 question-clustered 95% intervals	Robustness (D16)	Borderline flag
E3 UAE (and Singapore, secondary)	Registered	Codes per §4 and S1–S4
E3 "offered as an option" count	Post-registration descriptive (D5)	Not tabulated in the frozen analysis
E3 strict sensitivity	Robustness (D16)	—

Endpoint, analysis or readout	Status	Notes
E3b	Post-registration, exploratory (D6)	No figure in the frozen outputs
E4 UAE	Registered endpoint; quarantined (D17)	Only hand-checked counts reportable
E5 Singapore protocol first-line count	Registered	general + high-risk
E5 "first-line with metformin"	Post-registration, reported not counted (D2)	—
E5 strict sensitivity	Robustness (D16)	—
Saudi E3 (10-agent set)	Post-registration (D10, D12)	Every code provisional (status not established)
Saudi E3 opportunity denominator	Robustness (D16)	—
Saudi E3-i	Post-registration, exploratory (D12); denominator (D16)	Never a congress effect
Saudi non-programme context counts	Post-registration descriptive (T8)	Never coded
Blind second coding pass	Robustness (D16)	Same model family; not external validation
\$5 descriptive source analyses: (a) the Note 04 clinic page per cycle; (b) new local or regional pages drawing ≥ 10% of one engine's answers; (c) source-class mix	Registered (§5), reported late (D24)	§11.5.1
Band-rule readings; cycle-level change; per-cycle view; corrected self-check; brand-only naming; search rate; attribution sensitivity and minimum attainable p; Saudi E3-i use context; per-answer outcome table; E1 latency intervals and text-route miss; the local page behind Claude's UAE naming	Post-registration robustness (D24)	§11.5.2–§11.5.12; never replaces a registered verdict
Bridge offsets, negative control	Registered design (§6); execution changes D11, D15	—
Readouts: REIMAGINE 1, TRIUMPH-2, CagriSema appetite abstract 122, ACHIEVE programme, SMARTTEST, ACHIEVE-4 (LBA 25), proteomics (LBA 29)	Registered	—
Readouts: OCTANE (LBA 59)	Post-registration (D3)	—
Readouts: Saudi programme (agent set + teplizumab abstracts)	Post-registration (D10)	—

4. Data collection

4.1 Engines and models as recorded

Every one of the 67 pulled cycles records the same four models in its `providerModels`: `gpt-4o-mini`, `claude-haiku-4-5`, `sonar`. Captures in the pull, by provider and model: `openai:gpt-4o-mini` 3,019; `anthropic:claude-haiku-4-5` 3,016; `perplexity:sonar` 2,983; `google:gpt-4o-mini` 3,015. Admission (`lib.ts admit`) rejects a cycle whose recorded model or any capture's model differs from these. No cycle was rejected on that ground.

Endpoints as configured. These are developer API endpoints, not the consumer apps (`src/engine/providers.ts`; per-API citation definitions in §11.5.7):

- **ChatGPT:** `gpt-4o-mini` through the OpenAI Responses API with the built-in `web_search` tool. The model decides per request whether to search.
- **Claude:** `claude-haiku-4-5` through the Anthropic Messages API with the `web_search_20250305` server tool, at most 5 searches per request. The model decides whether to search.
- **Gemini:** `gemini-2.5-flash` through the Gemini API (`generateContent`) with Google Search grounding. Grounding is model-decided.
- **Perplexity:** `sonar` through the Perplexity chat-completions API, which retrieves on every request.

PN1 (post-publication). For every scheduled study cycle and panel run, AlphaCitation's request was the same on both sides of the 7 October change: the instrument's code sent model `sonar` with the same parameters, the `v5` setting in place since 27 September (§5.1), to the chat-completions endpoint; the bridges' `v4`-setting runs (5 October, UAE; 8 October, Singapore and Saudi Arabia) omitted that setting by design (§5.6); and the stored model string is `sonar` on every capture. The per-request parameters were not stored. Two changes in Perplexity's responses were found after publication. (1) Between 6 October 16:34 and 7 October 13:32 UTC the structure of the stored Perplexity answers changed at once, in this study and in a fixed 12-category monitoring panel. On each day from 7 to 13 October at least 98% of them carried exactly 10 search results (the lowest for one series on one day: 21 of 26, Singapore, 10 October), against at most 2% on each day from 20 September to 6 October (a median of 19–20 before). Answers became longer, with section headings, and carried fewer in-text citation markers (paper §6.2). (2) Before 28 September, 11 of the 36 cycles in which every other engine answered every question lacked between 1 and 6 Perplexity answers (30 in all); from 28 September none did. The failed calls were not stored, so their kind is unknown, and no cycle was excluded on this account (§4.6). The response identifier, the model field returned by the engine and the usage block were not recorded, so neither change can be attributed to a cause from the stored records.

No user-location parameter is sent on any of the four calls. The market is supplied only by the market-localised framing instruction that every question carries (wording withheld, §4.2). Searches therefore run from each API's default location, which is not what a user in the market would see.

The pull does not store search invocation itself. Citation presence is a lower bound on searching (§11.5.7).

4.2 Markets, measured medicines, batteries

Series	Measured medicine	Market frame	Questions per cycle	Governed (scored) questions per engine	Registration
uae-glp1	semaglutide (Wegovy)	UAE, persona general	60	54	Registered
singapore-sgt2	empagliflozin (Jardiance)	Singapore, persona general	26	19	Registered
saudi-t1d	teplizumab (Tzield)	Saudi Arabia; market-localised framing instruction, wording withheld (D13)	56	55	Post-registration (D10)

Governed population. Questions that ask about the measured medicine directly, or set it head to head with a named competitor ("focused" questions), are excluded from the naming denominator. This matches the stored governed counts. The rule is `promptPopulation.ts isFocusedPrompt`, applied to the prompt slot. Focused slots carry the opaque identifiers `*-fnn` in the evidence package.

This explains why the per-engine E2 denominators are 54, 19 and 55 per cycle, while E1, E3, E4 and E5 count every captured answer (60, 26 and 56 per engine per cycle).

Question slots. The study repeatedly sampled a maintained battery of question slots. Some slot texts differed across cycles; analyses requiring identical text used an explicitly matched subset. Prompt matching, exclusions and their denominators are reported separately from slot identity (§4.4, §4.5).

Exact question wording is not published (the amended disclosure boundary, §2.1). They are held in the local pull, and committed outputs carry `prompt_id` only. The question-instrument statement:

Question-instrument availability. The exact question battery is withheld to preserve the integrity of AlphaCitation’s longitudinal measurement instrument. Public disclosure of the active questions could enable content to be optimized specifically against the instrument, potentially influencing future observations and compromising comparability over time.

To support scientific scrutiny, the study reports its question taxonomy, sampling design, endpoint definitions, coding procedures, analytical methods and reproducibility controls. The complete instrument is preserved in the frozen study record.

This restriction represents a departure from CHART Item 5b and limits independent replication of the exact data-collection procedure.

What the supplement does disclose about the battery:

- question counts per series (above);
- the governed-population rule;
- the which-medicine and opportunity rules (§7.3, §7.5);
- the admission and prompt-match rules (§4.4).

The scorer, its brand-name construction and its calibration are production internals and are not described beyond the deterministic consistency check (§9.1).

4.3 Cycle schedule

- **Recorded completion times** (`scanned_at`, median per series over the admitted cycles, `easd-2026-analysis.json` `perCycleNaming`):
 - UAE: 14:05 UTC (22 cycles; the first, 22 Aug, at 17:44).
 - Singapore: 14:03 UTC (23 cycles; the first, 22 Aug, at 17:42).
 - Saudi Arabia: 14:05 UTC before the congress (11 cycles; the first, 24 Aug, at 19:40), and 16:34 UTC from 4 Oct (8 cycles).
- **Schedule behind those times:** the UAE and Singapore series run from the production cron at 14:00 UTC. From 4 Oct the Saudi series ran from a second scheduled job at 16:30 UTC (D20). A cycle completes a few minutes after its start.
- **Daily cadence:** 5–11 Oct for all three series (D9, D10).
- **v5 cycles before the congress.** One admitted pre-congress v5 cycle exists, in the UAE (27 Sep). Singapore has none: its 27 Sep cycle failed the capture rule. Saudi Arabia has none: its 27 Sep cycle captured 9 of 56 questions (§4.6).
- **Before 5 Oct:** admitted cycles are spaced irregularly, 2 to 8 days apart (dates in §10.1).
- **Window assignment** (`lib.ts windowOf`): a cycle’s start is its earliest capture time and its completion is `scanned_at`.
 - before-v4 or before-v5: completed before 28 Sep 06:00, by engine version.
 - during: started by 2 Oct 22:00.
 - after: completed by 11 Oct 23:59:59.
 - post-cutoff: anything later.

Admitted cycles per window, from `admissions` in the two final coding files:

Series	before-v4	before-v5	during	after	Total admitted
uae-glp1	13	1	1	7	22
singapore-sgl2	15	0	1	7	23
saudi-t1d	11	0	0	8	19

64 admitted of 67 pulled. Admitted answers (all captures): UAE 5,262; Singapore 2,386; Saudi 4,235. Sources: `answersPerEngineCycle` sums; the Saudi figure is also stated in the Saudi ledger §4.

4.4 Admission rules as applied

Series	Capture rule	Prompt-match rule	Cache rule	Source of the threshold
uae-glp1	≥ 95% of prompts captured	≥ 57 of 60 match the first cycle	0 cache hits	§2
singapore-sgl2	≥ 95%	26 of 26	0	§2
saudi-t1d	≥ 95%	≥ 52 of 56	0	D13, correcting D12’s 54 of 56

4.5 Every cycle and its admission

Source: `admissions` in `docs/evidence/computed/easd-2026-final-coding.json` and `docs/evidence/computed/easd-2026-saudi-final-coding.json`.

Series	Cycle	Date	Window	v	Prompts captured	Match first cycle	Captures	Admitted	Reasons
uae-glp1	UAE-C01	2026-08-22	before-v4	4	60/60	60	234	yes	—
uae-glp1	UAE-C02	2026-08-24	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C03	2026-08-26	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C04	2026-08-28	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C05	2026-08-30	before-v4	4	60/60	59	239	yes	—
uae-glp1	UAE-C06	2026-09-04	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C07	2026-09-09	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C08	2026-09-11	before-v4	4	60/60	59	240	yes	—
uae-glp1	UAE-C09	2026-09-14	before-v4	4	60/60	59	236	yes	—
uae-glp1	UAE-C10	2026-09-16	before-v4	4	60/60	57	236	yes	—

Series	Cycle	Date	Window	v	Prompts captured	Match first cycle	Captures	Admitted	Reasons
uae-glp1	UAE-C11	2026-09-21	before-v4	4	60/60	59	239	yes	—
uae-glp1	UAE-C12	2026-09-23	before-v4	4	60/60	58	240	yes	—
uae-glp1	UAE-C13	2026-09-25	before-v4	4	60/60	58	240	yes	—
uae-glp1	UAE-C14	2026-09-27	before-v5	5	60/60	58	238	yes	—
uae-glp1	UAE-C15	2026-09-29	during	5	60/60	58	240	yes	—
uae-glp1	UAE-C16	2026-10-04	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C17	2026-10-05	after	5	15/60	15	60	no	captures for 15 of 60 prompts (< 95%); 15 prompts match the first cycle (< 57)
uae-glp1	UAE-C18	2026-10-06	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C19	2026-10-07	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C20	2026-10-08	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C21	2026-10-09	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C22	2026-10-10	after	5	60/60	58	240	yes	—
uae-glp1	UAE-C23	2026-10-11	after	5	60/60	58	240	yes	—
singapore-sglt2	SG-C01	2026-08-22	before-v4	4	26/26	26	100	yes	—
singapore-sglt2	SG-C02	2026-08-24	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C03	2026-08-26	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C04	2026-08-28	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C05	2026-08-30	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C06	2026-09-04	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C07	2026-09-06	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C08	2026-09-08	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C09	2026-09-10	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C10	2026-09-13	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C11	2026-09-15	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C12	2026-09-17	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C13	2026-09-19	before-v4	4	26/26	26	103	yes	—
singapore-sglt2	SG-C14	2026-09-21	before-v4	4	26/26	26	103	yes	—
singapore-sglt2	SG-C15	2026-09-26	before-v4	4	26/26	26	104	yes	—
singapore-sglt2	SG-C16	2026-09-27	before-v5	5	17/26	17	68	no	captures for 17 of 26 prompts (< 95%); 17 prompts match the first cycle (< 26)
singapore-sglt2	SG-C17	2026-10-02	during	5	26/26	26	104	yes	—
singapore-sglt2	SG-C18	2026-10-05	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C19	2026-10-06	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C20	2026-10-07	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C21	2026-10-08	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C22	2026-10-09	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C23	2026-10-10	after	5	26/26	26	104	yes	—
singapore-sglt2	SG-C24	2026-10-11	after	5	26/26	26	104	yes	—
saudi-t1d	SA-C01	2026-08-24	before-v4	4	56/56	56	224	yes	—
saudi-t1d	SA-C02	2026-08-27	before-v4	4	56/56	52	224	yes	—
saudi-t1d	SA-C03	2026-08-30	before-v4	4	56/56	52	221	yes	—
saudi-t1d	SA-C04	2026-09-01	before-v4	4	56/56	52	224	yes	—
saudi-t1d	SA-C05	2026-09-03	before-v4	4	56/56	52	224	yes	—
saudi-t1d	SA-C06	2026-09-08	before-v4	4	56/56	52	223	yes	—
saudi-t1d	SA-C07	2026-09-10	before-v4	4	56/56	52	224	yes	—
saudi-t1d	SA-C08	2026-09-14	before-v4	4	56/56	52	223	yes	—
saudi-t1d	SA-C09	2026-09-16	before-v4	4	56/56	52	220	yes	—
saudi-t1d	SA-C10	2026-09-21	before-v4	4	56/56	52	219	yes	—
saudi-t1d	SA-C11	2026-09-26	before-v4	4	56/56	52	220	yes	—
saudi-t1d	SA-C12	2026-09-27	before-v5	5	9/56	5	22	no	captures for 9 of 56 prompts (< 95%); 5 prompts match the first cycle (< 52)
saudi-t1d	SA-C13	2026-10-04	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C14	2026-10-05	after	5	56/56	52	221	yes	—
saudi-t1d	SA-C15	2026-10-06	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C16	2026-10-07	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C17	2026-10-08	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C18	2026-10-09	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C19	2026-10-10	after	5	56/56	52	224	yes	—
saudi-t1d	SA-C20	2026-10-11	after	5	56/56	52	224	yes	—

4.6 Excluded cycles and why

Cycle	Window	Recorded reasons	Context
SG-C16 (Singapore, 27 Sep, v5)	before-v5	captures for 17 of 26 prompts (< 95%); 17 prompts match the first cycle (< 26)	Registered exclusion (§3). Singapore therefore has no admitted before-v5 cycle

Cycle	Window	Recorded reasons	Context
SA-C12 (Saudi, 27 Sep, v5)	before-v5	captures for 9 of 56 prompts (< 95%); 5 prompts match the first cycle (< 52)	D13. Saudi has no admitted pre-congress v5 cycle, so its v4→v5 comparison rests on the bridge alone
UAE-C17 (UAE, 5 Oct, v5)	after	captures for 15 of 60 prompts (< 95%); 15 prompts match the first cycle (< 57)	The cache incident (D14)

The cache incident (D14). The production scan reused an engine's answer to an identical brand-neutral question for the lifetime of a category cache, which was 24 hours. The daily cron fires at the same minute each day, so consecutive daily scans start a few minutes under 24 hours apart.

- The 5 Oct UAE cycle ran 23h56m after the 4 Oct cycle. It took 45 of its 60 questions per engine from the 4 Oct answers; only 15 per engine were fresh calls.
- Only the fresh calls were captured. That is why the admission record shows 15 prompts captured and 0 cache hits on capture rows: the cached answers left no capture. The cycle therefore failed the capture rule, which is the mechanism of D14's exclusion (D21 item 6).
- The cycle fails §2 and was excluded, never repaired.
- **Fix, deployed 5 Oct before the next daily cycle:** cache lifetime 12 hours, below the shortest allowed gap between two scans of a daily brand. A scheduler test fails if it reaches that gap.
- A second deploy, completed by 14:45 UTC on 5 Oct, moved the cache to a new key namespace, so 24-hour entries written on 5 Oct were unreachable on 6 Oct.
- The incident was found from capture counts and timestamps; no answer text was read.
- The bridge calls engines directly and never used this cache.
- **Effect:** the UAE after window has seven admitted cycles (4 Oct and 6–11 Oct). Singapore and Saudi Arabia were unaffected.

4.7 Cutoff and completeness of the final pull

The run took place at about 16:57 UTC; D20's text says about 17:15 UTC, and D21 item 4 corrects it. D20 records that the window's last scheduled cycle was Saudi, at 16:30 UTC on 11 Oct; it completed at 16:34. No study cycle was scheduled again before 23:59. A fresh pull at the time of the run returned the same 67 cycles as the pull of 16:36 UTC. The E1 manifest records its own pull as 2026-10-11T16:36:26.981Z, whose cycles are identical (D21 item 4); the analysis, coding and E1 tally record 2026-10-11T16:56:27.066Z.

5. Instrument change and bridge

5.1 The change

On 27 Sep the instrument moved from ENGINE_VERSION 4 to 5. A sampling setting was pinned on three engine paths (Claude, Gemini, Perplexity); the ChatGPT path was already pinned and did not change. Models, questions and scoring did not change. The parameter and its values are withheld (§10).

5.2 Construction

Source: docs/evidence/computed/easd-2026-bridge.json (design) and scripts/easd-2026-study/bridge.ts.

Battery and replicates.

- Each series' battery is the stored question texts of its latest admitted v5 cycle with a complete capture set.
- It is asked of the same four models in six interleaved runs on one day, in the order v4, v5, v5, v4, v4, v5 (A-B-B-A-A-B). That is R = 3 per arm.
- The v4 arm sends each vendor's default, exactly as the v4 request path did. The v5 arm sends the production pin.

Scoring and population.

- Every answer is scored by the production scorer, unmodified.
- The population is the governed one (focused questions excluded).

Estimand. For each engine and each question:

1. Compute the naming rate within each arm over its three runs.
2. Take the paired difference, v5 minus v4.
3. The offset is the mean of the paired differences over questions.
4. The 95% interval is a percentile bootstrap over questions: 2,000 resamples, mulberry32 seeded 20261005. The lower bound is the 51st order statistic; the upper bound is order statistic $\text{floor}(\{0.975 \times 2000\})$, which is the 1,950th.

The four-engine "all" offset pools the engine × question differences (54 × 4 = 216 for the UAE).

Validation before any paid call. The scorer re-scored the stored answers of the source cycle and had to reproduce its governed per-engine counts exactly, or stop.

Durability (D11). Every outcome was appended to a checkpoint as it arrived (run, arm, engine, question id, named or failed; no text). A failed call was retried; a scored call was never redrawn. The estimand is computed only from a complete set of six runs.

Negative control. The ChatGPT interval must contain 0, or the bridge is reported failed for that series and not applied.

5.3 Execution record

Series	Registered	Source cycle	Source completed	Ran on	First outcome	Last outcome	Questions	Failed calls	Validation (rescored = governed)	Rescored source counts (ChatGPT, Claude, Gemini, Perplexity)	Recorded cost (USD)
uae-glp1	yes	UAE-C16	2026-10-04T14:05:03.93548+00:00	2026-10-05	2026-10-05T12:44:30.616Z	2026-10-05T13:22:22.603Z	54	0	match	13/54, 36/54, 31/54, 23/54	35.79
singapore-sglt2	yes	SG-C20	2026-10-07T14:03:10.133859+00:00	2026-10-08	2026-10-08T13:16:08.018Z	2026-10-08T13:27:21.941Z	19	0	match	19/19, 19/19, 19/19, 19/19	14.22
saudi-t1d	no (D10)	SA-C16	2026-10-07T16:34:38.954606+00:00	2026-10-08	2026-10-08T14:11:15.608Z	2026-10-08T14:56:53.913Z	55	1	match	9/55, 6/55, 6/55, 1/55	36.17

5.4 Per-run naming counts (run-to-run spread)

uae-glp1

Run	Arm	ChatGPT	Claude	Gemini	Perplexity
1	v4	17/54	37/54	31/54	19/54
2	v5	16/54	36/54	31/54	18/54
3	v5	15/54	38/54	26/54	15/54
4	v4	16/54	36/54	32/54	16/54
5	v4	15/54	33/54	30/54	16/54
6	v5	14/54	37/54	29/54	18/54

singapore-sglt2

Run	Arm	ChatGPT	Claude	Gemini	Perplexity
1	v4	19/19	19/19	19/19	19/19
2	v5	19/19	19/19	19/19	19/19
3	v5	19/19	19/19	19/19	19/19
4	v4	19/19	19/19	18/19	19/19
5	v4	19/19	19/19	19/19	19/19
6	v5	19/19	19/19	19/19	19/19

saudi-t1d

Run	Arm	ChatGPT	Claude	Gemini	Perplexity
1	v4	11/55	0/55	5/55	0/55
2	v5	10/55	1/55	9/55	0/55
3	v5	7/55	2/55	2/54	1/55
4	v4	9/55	1/55	3/55	1/55
5	v4	8/55	2/55	3/55	0/55
6	v5	11/55	1/55	8/55	1/55

5.5 Offsets (v5 – v4) with 95% intervals

Values are proportions as stored (multiply by 100 for percentage points).

Series	Engine	Paired questions	Offset (v5 – v4)	95% interval	Interval contains 0
uae-glp1	ChatGPT (negative control)	54	-0.0185	[-0.0617, 0.0247]	yes
uae-glp1	Claude	54	0.0309	[-0.0617, 0.1173]	yes
uae-glp1	Gemini	54	-0.0432	[-0.1296, 0.0432]	yes
uae-glp1	Perplexity	54	0	[-0.0432, 0.0556]	yes
uae-glp1	Four engines	216	-0.0077	[-0.0432, 0.0278]	yes
singapore-sglt2	ChatGPT (negative control)	19	0	[0, 0]	yes
singapore-sglt2	Claude	19	0	[0, 0]	yes
singapore-sglt2	Gemini	19	0.0175	[0, 0.0526]	yes
singapore-sglt2	Perplexity	19	0	[0, 0]	yes
singapore-sglt2	Four engines	76	0.0044	[0, 0.0132]	yes
saudi-t1d	ChatGPT (negative control)	55	0	[-0.0424, 0.0424]	yes
saudi-t1d	Claude	55	0.0061	[-0.0121, 0.0242]	yes
saudi-t1d	Gemini	55	0.0485	[-0.0061, 0.103]	yes
saudi-t1d	Perplexity	55	0.0061	[0, 0.0182]	yes
saudi-t1d	Four engines	220	0.0152	[-0.0045, 0.0333]	yes

Negative control. It holds in all three series (`negativeControl.holds: true`). The stored verdict, "VALID — the ChatGPT control is consistent with zero; offsets may be applied", is a decision recorded by `bridge.ts`, not an independent validation. The unchanged-path control showed no detected shift within its estimated interval. No equivalence margin was prespecified. The bridge estimates a same-day setting difference and assumes that this offset applies to the before/after comparison; transport across dates and prompt contexts was not independently established.

Singapore. Naming is near ceiling: 19/19 on every engine in every run except Gemini run 4 (18/19). Its offsets can only show the absence of a change that a ceiling could hide. No Singapore band is bridged (I4).

Cost.

Item	Cost	Source
UAE bridge	\$35.79	meteredCostUsd
Lost first UAE attempt	about \$24, nothing used	D11
Singapore bridge	\$14.22	meteredCostUsd
Saudi bridge	\$36.17 recorded; D15 gives about \$41	see note

The Saudi figures differ (D21 item 5) because `bridge.ts` records the spend of the process that completes the series. The Saudi process stopped once after 171 calls (a file-sync lock on the local spend counter) and resumed from its checkpoint the same day. The recorded \$36.17 therefore covers the resumed process only. D15's "about \$41" is the series total.

5.6 The registered bridge self-check (§6; reported under D23)

§6 registered that the v4-setting run-to-run spread "should reproduce the v4 band's width, a check on the bridge itself", with the v5 spread reported beside it. No threshold was registered, so the comparison is shown, not scored.

The analysis did not report this check until D23 (commit 66875044), after an internal audit identified the omission. It is computed from `bridge.json` and stored in `easd-2026-analysis.json` (`bridgeSelfCheck`). The table is copied verbatim from `docs/evidence/easd-2026-analysis.md`. Since commit 73cfbdc3, the band column is labelled by source: UAE from the registered \$4 band, Saudi from I5, and Singapore as its v4 range, which is not a registered band and is used in no decision (I4).

Series	Engine	v4 runs (K/N)	v4 same-day spread	v5 runs (K/N)	v5 same-day spread	Reference band (UAE: registered \$4; Saudi: I5; Singapore: v4 range, no registered band)	Band width	Spread / width	Reproduces width
uae-glpl	Perplexity	19/54, 16/54, 16/54	29.6%–35.2% (5.56pp)	18/54, 15/54, 18/54	27.8%–33.3% (5.56pp)	46.3%–58.8%	12.5pp	0.445	no
uae-glpl	Gemini	31/54, 32/54, 30/54	55.6%–59.3% (3.7pp)	31/54, 26/54, 29/54	48.1%–57.4% (9.26pp)	48.1%–74.1%	26pp	0.142	no
uae-glpl	ChatGPT	17/54, 16/54, 15/54	27.8%–31.5% (3.7pp)	16/54, 15/54, 14/54	25.9%–29.6% (3.7pp)	20.4%–44.4%	24pp	0.154	no
uae-glpl	Claude	37/54, 36/54, 33/54	61.1%–68.5% (7.41pp)	36/54, 38/54, 37/54	66.7%–70.4% (3.7pp)	40.7%–55.6%	14.9pp	0.497	no
singapore-sglt2	Perplexity	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	100.0%–100.0%	0pp	n/a	n/a (ceiling)
singapore-sglt2	Gemini	19/19, 18/19, 19/19	94.7%–100.0% (5.26pp)	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	94.7%–100.0%	5.26pp	1	n/a (ceiling)
singapore-sglt2	ChatGPT	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	100.0%–100.0%	0pp	n/a	n/a (ceiling)
singapore-sglt2	Claude	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	19/19, 19/19, 19/19	100.0%–100.0% (0pp)	94.7%–100.0%	5.26pp	0	n/a (ceiling)
saudi-t1d	Perplexity	0/55, 1/55, 0/55	0.0%–1.8% (1.82pp)	0/55, 1/55, 1/55	0.0%–1.8% (1.82pp)	3.8%–21.8%	18.04pp	0.101	no
saudi-t1d	Gemini	5/55, 3/55, 3/55	5.5%–9.1% (3.64pp)	9/55, 2/54, 8/55	3.7%–16.4% (12.66pp)	3.6%–20.0%	16.36pp	0.222	no
saudi-t1d	ChatGPT	11/55, 9/55, 8/55	14.5%–20.0% (5.45pp)	10/55, 7/55, 11/55	12.7%–20.0% (7.27pp)	10.9%–30.9%	20pp	0.273	no
saudi-t1d	Claude	0/55, 1/55, 2/55	0.0%–3.6% (3.64pp)	1/55, 2/55, 1/55	1.8%–3.6% (1.82pp)	0.0%–10.9%	10.91pp	0.334	no

Result. In every engine of the UAE and Saudi series, the three same-day v4-setting runs spread over 0.101 to 0.497 of the band's width. "Reproduces width" is "no" in all eight rows. Singapore is at ceiling. The check therefore does not hold as worded.

Interpretation (D23), as corrected by D24. D23 read the table as showing that between-day variation dominates within-day variation. That reading did not allow for unequal numbers of runs: a range over 3 runs is expected to be narrower than a range over 7 to 13 cycles even with no day effect.

The corrected self-check (D24, §11.5.5) divides each range by d2(n) and compares SDs. On the v4 runs:

- **No material excess between days:** UAE Claude (d2-adjusted ratio 0.985) and UAE Perplexity (0.708).
- **Material excess:** UAE Gemini, UAE ChatGPT, Saudi Perplexity and Saudi Gemini.
- **One-sided F test** (4 df within), $p < 0.05$ for UAE ChatGPT, Saudi Perplexity and Saudi Claude.

PN1 (post-publication). The Saudi and Singapore bridges ran on 8 October, after the change in Perplexity's responses (paper §6.2). Saudi Perplexity's v4-setting runs on that day (0, 1 and 0 of 55) and its v4 band (2 of 53 to 12 of 55 per cycle) therefore come from different response regimes, so its "material excess" classification and its F test do not isolate between-day variation.

The 0.5 threshold for "material" was chosen after the ratios were seen, so that classification is descriptive. The within-day SDs rest on 2 (or 4 pooled) degrees of freedom and are very imprecise. The v5 same-day spreads run from 1.82 to 12.66 points; the largest is Saudi Gemini, whose band is 16.36 points wide.

So between-day variation dominates for some engines and not for others. For UAE Claude, the engine that carries the one declared movement, the same-day spread is what the run count alone predicts. What still holds:

- The paired, same-day bridge measures the setting offset it was designed to measure.
- The band, built over weeks, remains the yardstick for between-cycle comparison.
- No registered verdict changes.

A descriptive observation, not a registered analysis. D23 notes that the same-day v4-setting runs on 5 Oct place Claude at 61.1%–68.5% (106 of 162 pooled). That is above its v4 band, before the instrument offset is applied.

Before the congress, observed naming rates varied across cycles; this study did not determine whether contemporaneous evidence or service changes contributed to that variation.

5.7 Exact-lattice restoration of the offset bounds (D19)

A paired mean over Q questions × 3 runs per arm moves in steps of 1/(3Q). The stored bounds are rounded to four decimals. D19 restores each stored bound to the nearest lattice value when it lies within 0.00005 of it (`analyze.ts exactBound`):

Series	Paired questions	Lattice step
UAE, per engine	54	1/162
UAE, four engines	216	1/648
Saudi, per engine	55	1/165
Saudi, four engines	220	1/660

The registered decision uses the stored bounds. The restored bounds feed only the full-precision columns (§6.4).

6. Baseline bands

6.1 Registered UAE bands

The registered bands are the v4 single-cycle ranges quoted in §4: Perplexity post-step 46.3–58.8%, Gemini 48.1–74.1%, ChatGPT 20.4–44.4%, Claude 40.7–55.6%, four engines 41.2–51.2%. "Post-step" means Perplexity's band uses only the v4 cycles from 9 Sep, as in Note 04 (analyze.ts UAE_PERPLEXITY_STEP_FROM). Every other engine uses all 13 admitted v4 cycles.

6.2 Saudi bands (I5)

Each engine's Saudi band is the arm's own v4 single-cycle range over its 11 admitted pre-congress cycles, computed from the governed counts at full precision.

6.3 How a bridged band is formed (I1)

Bridged band (§4 E2): lower end = registered v4 low + the lower bound of the engine's v5–v4 offset interval; upper end = registered v4 high + the upper bound. A movement is declared when the after-window pooled rate lies outside it.

The pre-congress reference for the magnitude label (I2) is the v4 pooled rate over the cycles that define the band, plus the engine's offset. The four-engine band uses each cycle's four-engine pooled rate.

Sensitivity to the reading (D24, §11.5.2). The registered words "shifted by the ... offset ... with each end widened by the offset's 95% interval" admit other readings than I1. I1 was fixed in D16, after per-cycle counts had been displayed.

- Shifting by the point offset and widening each end by the interval's half-width makes Perplexity (UAE) a movement (+0.546pp).
- Widening by the full width makes Claude (UAE) no movement (–8.865pp).
- No other verdict changes under either reading.
- The registered verdicts are those of I1 and stand.

Per-cycle comparison uses the bridged band. Protocol §2: "No during- or after-window cycle is ever compared with a v4 figure except through §6". Every per-cycle view in this supplement (§11.5.4) therefore compares each v5 cycle with the bridged band, never with the raw v4 band. The registered decision itself pools the after window.

6.4 Full-precision band table

Source: e2[].band, e2[].precision and e2[].after in docs/evidence/computed/easd-2026-analysis.json. Rates are shown as percentages to three decimals; the JSON holds full precision.

Distance is signed: positive means beyond the nearer edge, negative means inside it. "Registered" uses the band as written with the stored offset bounds. "Exact cycle" uses the band rebuilt from exact single-cycle rates with lattice-restored bounds.

Series	Engine	Registered band (as written)	Exact cycle band	Offset 95% (stored)	Bridged band, as decided	Bridged, exact registered band	Bridged, exact cycle band	After K/N	After rate (exact)	Distance, registered (pp)	Distance, exact registered (pp)	Distance, exact cycle (pp)	Clustered 95% (robustness)	Nearer edge	Clustered interval contains it	Verdict (registered)
uae-glpl	Perplexity	46.300%–58.800%	46.296%–58.824%	[–0.0432, 0.0556]	41.980%–64.360%	41.979%–64.356%	41.975%–64.379%	243/378	64.286%	–0.0743	–0.0698	–0.0934	56.610%–71.960%	64.360%	yes (borderline)	no movement
uae-glpl	Gemini	48.100%–74.100%	48.148%–74.074%	[–0.1296, 0.0432]	35.140%–78.420%	35.137%–78.421%	35.185%–78.395%	206/378	54.497%	–19.3574	–19.3603	–19.3122	45.500%–63.760%	35.140%	no	no movement
uae-glpl	ChatGPT	20.400%–44.400%	20.370%–44.444%	[–0.0617, 0.0247]	14.230%–46.870%	14.227%–46.869%	14.198%–46.914%	110/378	29.101%	–14.8705	–14.8734	–14.903	18.780%–39.150%	14.230%	no	no movement
uae-glpl	Claude	40.700%–55.600%	40.741%–55.556%	[–0.0617, 0.1173]	34.530%–67.330%	34.527%–67.328%	34.568%–67.284%	256/378	67.725%	0.3949	0.3965	0.4409	57.140%–78.040%	67.330%	yes (borderline)	movement
uae-glpl	Four engines	41.200%–51.200%	41.204%–51.174%	[–0.0432, 0.0278]	36.880%–53.980%	36.879%–53.978%	36.883%–53.951%	815/1512	53.902%	–0.0779	–0.0757	–0.0494	47.160%–60.910%	53.980%	yes (borderline)	no movement
saudi-tld (post-reg.)	Perplexity	3.774%–21.818%	3.774%–21.818%	[0, 0.0182]	3.774%–23.638%	3.774%–23.636%	3.774%–23.636%	26/440	5.909%	–2.1355	–2.1355	–2.1355	2.950%–9.090%	3.774%	yes (borderline)	no movement
saudi-tld (post-reg.)	Gemini	3.636%–20.000%	3.636%–20.000%	[–0.0061, 0.103]	3.026%–30.300%	3.030%–30.303%	3.030%–30.303%	40/440	9.091%	–6.0645	–6.0606	–6.0606	4.320%–14.550%	3.026%	no	no movement
saudi-tld (post-reg.)	ChatGPT	10.909%–30.909%	10.909%–30.909%	[–0.0424, 0.0424]	6.669%–35.149%	6.667%–35.152%	6.667%–35.152%	81/440	18.409%	–11.74	–11.7424	–11.7424	9.770%–27.730%	6.669%	no	no movement
saudi-tld (post-reg.)	Claude	0.000%–10.909%	0.000%–10.909%	[–0.0121, 0.0242]	–1.210%–13.329%	–1.212%–13.333%	–1.212%–13.333%	28/437	6.407%	–6.9218	–6.926	–6.926	3.200%–10.300%	13.329%	no	no movement
saudi-tld (post-reg.)	Four engines	6.944%–20.000%	6.944%–20.000%	[–0.0045, 0.0333]	6.494%–23.330%	6.490%–23.333%	6.490%–23.333%	175/1757	9.960%	–3.4657	–3.4703	–3.4703	6.440%–13.940%	6.494%	yes (borderline)	no movement

The same table as the frozen analysis prints it (docs/evidence/easd-2026-analysis.md, "E2 at full precision"):

Series	Engine	After (K/N, exact)	Distance, registered band (pp)	Distance, exact cycle band (pp)	Registered verdict	Same verdict on exact cycle band	Clustered 95% contains the nearer edge
uae-glpl	Perplexity	243/378 = 64.286%	–0.074	–0.093	no movement	yes	yes (borderline)
uae-glpl	Gemini	206/378 = 54.497%	–19.357	–19.312	no movement	yes	no

Series	Engine	After (K/N, exact)	Distance, registered band (pp)	Distance, exact cycle band (pp)	Registered verdict	Same verdict on exact cycle band	Clustered 95% contains the nearer edge
uae-glp1	ChatGPT	110/378 = 29.101%	-14.870	-14.903	no movement	yes	no
uae-glp1	Claude	256/378 = 67.725%	+0.395	+0.441	movement	yes	yes (borderline)
uae-glp1	Four engines	815/1512 = 53.902%	-0.078	-0.049	no movement	yes	yes (borderline)
saudi-t1d (post-reg.)	Perplexity	26/440 = 5.909%	-2.135	-2.135	no movement	yes	yes (borderline)
saudi-t1d (post-reg.)	Gemini	40/440 = 9.091%	-6.064	-6.061	no movement	yes	no
saudi-t1d (post-reg.)	ChatGPT	81/440 = 18.409%	-11.740	-11.742	no movement	yes	no
saudi-t1d (post-reg.)	Claude	28/437 = 6.407%	-6.922	-6.926	no movement	yes	no
saudi-t1d (post-reg.)	Four engines	175/1757 = 9.960%	-3.466	-3.470	no movement	yes	yes (borderline)

7. Endpoint coding manuals

7.1 E1: congress penetration and latency

7.1.1 Units and routes

The unit is the answer. Per engine per cycle, E1 counts answers that qualify by either of two routes.

- **Source route.** The answer cites a URL entered in the hand ledger as congress-derived, class (a)–(d), for the same series.
- **Text route.** The answer's own text reports a result of a readout.

Readout ids and the presentation times used for latency (e1.json readouts):

Readout	Series	Presented (UTC)	Registration
REIMAGINE 1 (CagriSema), LBA 11/12	uae-glp1	2026-09-29T12:45:00Z	registered
TRIUMPH-2 (retatrutide)	uae-glp1	2026-09-30T06:30:00Z	registered
CagriSema appetite, abstract 122	uae-glp1	2026-09-30T12:30:00Z	registered
ACHIEVE programme (orforglipron)	uae-glp1	2026-10-01T06:30:00Z	registered
SMARTTEST (SGLT2 vs metformin first line)	singapore-sglit2	2026-10-01T07:30:00Z	registered
ACHIEVE-4 kidney outcomes, LBA 25	uae-glp1	2026-10-01T13:30:00Z	registered
Tirzepatide vs semaglutide proteomics, LBA 29	uae-glp1	2026-10-01T13:30:00Z	registered
OCTANE (switch to oral semaglutide, real-world), LBA 59	uae-glp1	2026-09-30T09:45:00Z	post-registration (D3)
EASD 2026 type 1 diabetes programme (agent set + teplizumab abstracts)	saudi-t1d	2026-09-29T08:00:00Z	post-registration (D10)
Other EASD 2026 data (not in the readout set)	uae-glp1, singapore-sglit2, saudi-t1d	2026-09-28T06:00:00Z	post-registration (E1 classes (c)/(d))

Two conventions in those times:

- **OCTANE** was verified against the EASD 2026 Final Programme on 11 Oct, which is the check D3 required (D21 item 3). The programme lists "LBA 59 3-month real-world weight outcomes among patients with overweight or obesity after switching from injectable semaglutide or tirzepatide to oral semaglutide for weight management: the OCTANE ..." in LBA Short Oral Event C (LBA SO 03), Wednesday 30 Sep, 11:45–12:45 CEST. The latency origin is the session start, 09:45 UTC. Before D21 it was a date-only midnight, which made the OCTANE latencies about 0.4 day longer. The check was made after the freeze. It remains a post-registration readout (D3).
- **"Other EASD 2026 data"** runs from the opening of the congress window.

Results public before the congress session. Sponsor topline releases put the primary results of two readout-set trials in the public domain before their EASD sessions:

- **TRIUMPH-2:** Lilly topline release of 23 Jul 2026 (Eli Lilly and Company press release, 23 July 2026; paper reference 18), confirmed by Drug Topics on 25 Jul 2026 (Gallagher, Drug Topics; paper reference 19).
- **ACHIEVE-4:** Lilly topline release of 16 Apr 2026 (Eli Lilly and Company press release, 16 April 2026; paper reference 20), confirmed by BioPharm International on 16 Apr 2026 (Schoenthaler, BioPharm International; paper reference 21). That release concerned the trial's primary cardiovascular result, not the registered kidney-outcome analysis (LBA 25).

The sponsor pages returned HTTP 403. The releases were therefore read in syndicated copies (recorded as verifiedBy in docs/research/easd-2026/references.json), with the trade reports beside them.

Classification of the readout set (readout classification record, literature-review.md §9, with ACHIEVE-4 separated into its kidney-outcome analysis and its primary cardiovascular result). This classification is used in place of the earlier phrase "first disclosed at EASD":

Class	Readouts
First disclosed at EASD as far as established	SMARTTEST; OCTANE (post-registration, D3); ACHIEVE-4 kidney-outcome analysis (LBA 25; earlier disclosure not established)
Full data at or around EASD, primary topline public earlier	TRIUMPH-2; ACHIEVE-4 primary cardiovascular result (not itself a registered readout)
Re-presented, or secondary analyses, at EASD	REIMAGINE 1; most of the ACHIEVE programme

The CagriSema appetite abstract (122) and the proteomics study (LBA 29) are not assigned a class in that list.

Consequences for E1.

- TRIUMPH-2's presentation-based detection interval concerns the full data rather than its earlier headline result.

- Earlier disclosure of the registered ACHIEVE-4 kidney analysis was not established. A cardiovascular topline must not be treated as prior disclosure of that kidney analysis.
- A topline source dated before 28 Sep fails the date rule of E1 classes (c) and (d), so it can never qualify, even if an answer cites it.
- Class (b) has no date condition (§7.1.2).
- The E1 ledger records the same point for TRIUMPH-2: "a topline release before the congress does not change it".

7.1.2 New-URL scope (D16, e1.ts)

A URL is new when no before-window cycle of the same series cited it, by any engine. Normalisation:

- `utm_*` parameters, trailing slashes and a leading `www.` are removed.
- Gemini's Google-hosted grounding redirect links are resolved by their `Location` header with `redirect: "manual"`: read, not followed. No redirect link is reproduced here (the publication's Gemini constraints, §13.3).

Gemini's pre-congress redirect links were not resolved. Pages earlier cited only through Gemini therefore count as new. This over-cludes pages for reading, which cannot create a qualifying source.

Detection limit. A page cited before the congress and later updated with congress content is not re-read. The text route flags an answer that reports a readout result only when the answer carries a congress or readout term (§7.1.5).

Class (b) carries no date condition (D21 item 8). Classes (c) and (d) require a date of 28 Sep or later. Class (b), "a journal publication of a readout-set trial", does not. A readout-set trial published before the congress would therefore qualify if newly cited in the congress or after window. D21 records that no readable new URL in any window is a readout-set publication, so the gap does not affect these results (§12, item 15).

7.1.3 Screen and hand ledger

Screen (e1.ts --stage screen).

- Every new (series, URL) pair in the during and after windows was fetched once: research user agent, 20-second timeout, 1.5 MB cap.
- A fetch that failed, or returned HTTP ≥ 400, was retried once with an ordinary browser user agent (--retry-unreadable).
- Title, printed page date and sentences carrying congress or readout terms were extracted.
- Pages with any such term became candidates for hand classification.

Scope of the inspection. All new series–URL pairs were subjected to an automated fetch and term screen. Candidate pages were classified from retrieved content. Of 5,268 pairs, 947 remained unreadable after retry. Both the source and text screens have incomplete and unquantified recall, including known misses of unnamed readout results (§7.1.5).

From e1.json screen:

New (series, URL) pairs	Unique URLs	Candidates read by hand	Screened out (no term)	Of which unreadable
5268	5266	114	5154	947

"Unreadable" means a fetch failure or HTTP ≥ 400 after the retry. An unreadable page carries no term, so the 947 sit inside the 5,154 screened out.

Final status of the 947, from docs/evidence/computed/easd-2026-e1-manifest.json `HttpStatus`:

Status	403	TypeError (network)	404	AbortError (timeout)	429	410	500	401	503	400	522	525	423
Pairs	619	170	77	28	19	14	5	4	4	2	2	2	1

Hand ledger (docs/evidence/easd-2026-e1-ledger.json).

- **Coder:** "Study coordinator (Claude Opus 5.5): every screened candidate page read and classified into prereg §4 E1 classes by its own article and date, never by page furniture added after citation; every text-route candidate read". Coded 2026-10-11.
- **Ledger scope:** "Cycles in the congress and after windows through 10 Oct 2026 (Saudi through 10 Oct). 11 Oct is added on 11 Oct."
- The ledger holds 122 source entries: 10 qualifying (class c-sponsor 3, class d-media 7; none in class a or b) and 112 excluded as not congress-derived (UAE 30, Singapore 17, Saudi 65).
- Eight entries were not in the script's candidate list. Each is a page the script could not read, read by hand in a browser, or added on review: the diaTribe round-up, a Tandem release, a Cureus article, a GoodRx page, a Mayo Clinic item and three brand pages. A page read by hand and entered in the ledger is no longer counted as unreadable (e1.ts, tally stage).
- **Text route:** every during- and after-window answer containing a congress or readout term was read. That gave 30 candidates; 0 qualified; uncoded text candidates 0.
- Each entry records the URL, class, readouts, printed page date, fetch date and one qualifying sentence. Contested entries also carry `contentAtCitation`, `readoutsUnverifiedAtCitation`, `adjudication` and `revisions`.

7.1.4 Latency: days to first qualifying detection

Label. Days from the specified presentation or programme anchor to first qualifying detection in an admitted sampled cycle.

Latency for engine × readout is measured from presentation to the completion time (`scanned_at`) of the first admitted cycle of the readout's registered series with a qualifying answer carrying that readout, in days rounded to 0.1. With no such answer, the cell reads "not observed by day N", where N is measured to the last admitted cycle of that series. A readout counts only in its registered series. A readout reaching another series is listed apart (D18).

Sampled detection brackets (D24, §11.5.11(a)). The last negative and first positive sampled cycles define observation brackets of [0, 4.2] days for Perplexity and [4.2, 6.2] days for Gemini (OCTANE). These brackets do not establish the engines' true evidence-uptake times, and the observation schedule and imperfect detection do not permit ranking the engines by uptake speed.

- Perplexity's bracket opens at 0 because no UAE cycle was admitted between 29 Sep and 4 Oct.
- Gemini's opens at 4.2 because the 5 Oct cycle was excluded.
- Daily cycles began only on 5 Oct (D9).

"Not observed by day N" runs to the completion of the series' last admitted cycle (UAE 14:05 UTC, Singapore 14:02 UTC and Saudi Arabia 16:34 UTC on 11 Oct), not to the 23:59 cutoff. A latency therefore measures the schedule as much as the engine.

Reading a latency. Detection is imperfect and may be non-persistent. "Not observed by day N" means not detected within follow-up, and refers to this battery and observation schedule. Programme-anchored exploratory rows do not give readout-specific uptake times.

7.1.5 The unverified rule, and unverified at citation (D19)

- **Unverified answer.** An answer that qualifies by neither route but cites at least one new URL that could not be read. It is counted as "+N unverified" beside the count and never as a negative observation.
- **Text-route recall is limited (D24, §11.5.11(b)).** The text route screens answers for congress and readout terms, so an answer that states a result without naming the trial is invisible to it. Two Gemini answers (6 and 7 Oct) state the OCTANE result (the figure matches reference value C-OCTANE-pct; the answers' wording is not reproduced, §13.3) without naming the trial; the E4 hand check found them. The 6 Oct answer qualifies through its cited source. The 7 Oct answer is not counted by E1 at all; the corrected either-route tally is in §10.2. Gemini's OCTANE latency is unchanged. "0 answers reported a readout result" therefore holds for answers that name a readout. The screen's recall for unnamed readout content elsewhere has not been estimated.
- **Unverified at citation (D19).** A readout that a qualifying source carries today but that cannot be shown to have been on the page when the engine cited it: no archived copy, and not echoed by the answer. It is reported, never counted in latency or E2 attribution, and never treated as negative.

7.1.6 Inspection coverage and unreadable-domain concentration

Source: inspectionCoverage.bySeriesEngineWindow in docs/evidence/computed/easd-2026-e1.json. "Read" is (new-URL citations – unreadable citations) / new-URL citations.

Series	Engine	Window	Answers	Citing a new URL	Citing an unreadable page	New-URL citations	Unreadable citations	New-URL citations read
saudi-t1d	Claude	after	445	212	72	371	85	77.1%
saudi-t1d	Gemini	after	448	435	412	5158	1626	68.5%
saudi-t1d	ChatGPT	after	448	119	32	226	38	83.2%
saudi-t1d	Perplexity	after	448	340	66	778	77	90.1%
singapore-sglt2	Claude	after	182	31	7	46	8	82.6%
singapore-sglt2	Claude	during	26	5	0	5	0	100.0%
singapore-sglt2	Gemini	after	182	158	107	1267	301	76.2%
singapore-sglt2	Gemini	during	26	24	15	194	44	77.3%
singapore-sglt2	ChatGPT	after	182	59	21	89	23	74.2%
singapore-sglt2	ChatGPT	during	26	1	1	1	1	0.0%
singapore-sglt2	Perplexity	after	182	73	27	308	33	89.3%
singapore-sglt2	Perplexity	during	26	19	7	132	10	92.4%
uae-glp1	Claude	after	420	261	76	628	101	83.9%
uae-glp1	Claude	during	60	24	6	52	7	86.5%
uae-glp1	Gemini	after	420	418	320	6006	768	87.2%
uae-glp1	Gemini	during	60	60	44	807	87	89.2%
uae-glp1	ChatGPT	after	420	197	54	378	58	84.7%
uae-glp1	ChatGPT	during	60	16	5	27	5	81.5%
uae-glp1	Perplexity	after	420	372	97	1038	129	87.6%
uae-glp1	Perplexity	during	60	40	8	103	21	79.6%

In the after window the share read runs from 68.5% (Saudi, Gemini) to 90.1% (Saudi, Perplexity). In the during window it runs from 0.0% (Singapore ChatGPT, one citation) to 100.0% (Singapore Claude, five citations).

Unreadable domains (top ten per series; counts are unreadable citations). The rows are generated from the evidence package's redaction-token-map.json (unreadableTopDomains, package v1.0). Every source the source-identification policy withholds appears as the package's token (§13.3), and the token types are the package's own.

Series	Top unreadable domains (count)
uae-glp1	[gemini-only host] (71), [gemini-only host] (50), [local-clinic-or-hospital-022] (48), marketresearch.com (44), qz.com (36), arabianbusiness.com (34), [commercial-or-unclassified-site-192] (31), goodrx.com (29), forbes.com (27), [gemini-only host] (26)
singapore-sglt2	mims.com (305), goodrx.com (18), [pharmacy-39] (13), dailypharm.com (11), journal-news.com (5), diatribe.org (5), nhathuocankhang.com (5), [clinic-or-hospital-055] (5), accessdata.fda.gov (4), [gemini-only host] (4)
saudi-t1d	[local-pharmacy-45] (412), [local-retailer-or-delivery-20] (298), [local-commercial-or-unclassified-site-120] (223), [local-pharmacy-33] (134), [local-retailer-or-delivery-05] (108), breakthrought1d.org.au (46), researchgate.net (44), diabetesjournals.org (43), [local-clinic-or-hospital-153] (31), indiamart.com (27)

The concentration is plain. In the Saudi after window, Gemini's 5,158 new-URL citations include 1,626 unreadable. The five largest Saudi unreadable entries (1,175 citations) are all tokenised: [local-pharmacy-45] (412), [local-retailer-or-delivery-20] (298), [local-commercial-or-unclassified-site-120] (223), [local-pharmacy-33] (134), [local-retailer-or-delivery-05] (108). In Singapore, mims.com, which the table names, accounts for 305 of the unreadable citations.

7.1.7 The diaTribe adjudication (D18, D19), in full

Source: revisedSources[0] in docs/evidence/computed/easd-2026-e1.json and the matching ledger entry, reproduced verbatim. Before D21 the adjudication quoted the Saudi question. D21 item 2 replaced the quotation at source with a description, as protocol §10 requires.

Current entry.

Field	Value
Series	saudi-t1d
URL	https://diatribe.org/diabetes-research/top-diabetes-news-easd-2026
As cited	direct (exact URL, not a redirect)
Engine	Claude
First observed in an answer	2026-10-08T16:31 (cycle SA-C17)
Class	d-media
Readouts	EASD 2026 type 1 diabetes programme (agent set + teplizumab abstracts); TRIUMPH-2 (retatrutide); Other EASD 2026 data (not in the readout set)

Field	Value
Readouts unverified at citation	TRIUMPH-2 (retatrutide)
Page date	2026-10-05
Fetches	2026-10-11 (the scripted fetch of 10 Oct returned HTTP 403, title "Just a moment...")

Revision (revisions[0], revised 2026-10-11; previous class d-media; previous readouts: Other EASD 2026 data (not in the readout set)). Reason, verbatim:

Full re-read of the article. Two of its items carry readouts without naming the trial, which the term screen cannot see. (1) "Retatrutide Promotes Weight Loss in People With Type 2 Diabetes": a late-stage trial of 1,152 people with type 2 diabetes and overweight or obesity over 80 weeks, nearly 19% weight loss at the highest dose and A1C -1.45%. These match the TRIUMPH-2 presentation of 30 Sep (1,152 adults, 80 weeks, 18.8% at 12 mg on the treatment-regimen estimand; Lilly's release gives 20.8% on the efficacy estimand). Class (d) turns on the article date (5 Oct) and its reporting congress data; a topline release before the congress does not change it. (2) "Tzield Therapy Begins in European Countries": first real-world teplizumab courses in stage 2 type 1 diabetes in the UK (10 people) and Italy (18 children), reported in the EASD 2026 round-up; teplizumab abstracts are in the Saudi programme readout (D10). The abstract number (373, 376 or 745) was not matched. The earlier entry recorded TRIUMPH-2 in its note but not in its readouts.

Content at citation (contentAtCitation), verbatim:

Cited by Claude, exact URL (not a redirect), in Saudi answer 87f5a541a5 of cycle SA-C17, captured 2026-10-08T16:31Z, three days after the article date (5 Oct). No archived copy exists (Wayback Machine: no snapshot, checked 11 Oct). The Tzield item is corroborated: the answer repeats it ("Real-world studies continue to provide evidence of its safety and efficacy as a therapy for slowing the progression of type 1 diabetes"; "temporary drops in white blood cell counts, rashes, and gastrointestinal issues"). The retatrutide item is not corroborated: the answer does not use it, and today's page cannot show it was there on 8 Oct.

Adjudication (adjudication), verbatim:

URL: <https://diatribe.org/diabetes-research/top-diabetes-news-easd-2026>, cited verbatim by Claude in one admitted Saudi answer (87f5a541a5, cycle SA-C17, 2026-10-08T16:31Z, a question asking for an evaluation of Tzield; question text withheld under protocol §10); cited in no UAE or Singapore answer. Registered eligibility: class (d), media dated 5 Oct (on or after 28 Sep) reporting congress data, so the answer counts as citing a congress-derived source. TRIUMPH-2: the article's retatrutide item satisfies the TRIUMPH-2 classification without naming the trial (same population, size, duration and effect as the 30 Sep presentation). Historical content: no archived copy; the article date precedes the answer by three days; the Tzield item is shown to have been present (the answer repeats it), the retatrutide item is not. Verdict: QUALIFYING SOURCE (class d) for the Saudi arm, labelled with the Saudi programme readout (teplizumab real-world item, verified at citation) and other EASD 2026 data. TRIUMPH-2 is UNVERIFIED AT CITATION: reported as unverified, not counted, and not treated as negative. It could never enter the registered TRIUMPH-2 cell, which belongs to the UAE series, where no answer cited this page: the registered UAE result stays "not observed by day 11.3".

What the shared wording shows. Shared wording between a captured answer and a later-fetched page corroborates textual overlap. Without a contemporaneous archive, it does not establish the page's complete historical content, the direction of copying or exclusive source provenance. The diaTribe Tzield item and the local page of §11.5.12 are read the same way.

Consequences in the tables.

- The Saudi Claude answer counts in E1 through the Saudi programme readout (latency 9.4) and "Other EASD 2026 data" (latency 10.4).
- TRIUMPH-2 appears only in the cross-arm table, as unverified (Claude, Saudi, day 8.4).
- The registered UAE TRIUMPH-2 cell stays "not observed by day 11.3".
- The page was cited in no UAE or Singapore answer.
- In E2, the Saudi Claude answer is a focused question and lies outside the governed population. The Saudi Claude attribution row therefore shows 0 E1 answers (§7.2.3).

The external check, and the other multi-topic sources. An external reviewer's independent check flagged the omission (D18). The correction kept the earlier label in revisions. The two other multi-topic qualifying sources were re-read in full under D18:

- BioPharma Dive mentions retatrutide only as data from May, before the congress.
- Patient Care's CagriSema terms sit in a related-article item, not the article.

7.1.8 Gemini, OCTANE and the PR Newswire copy

From 5 Oct, Gemini cited a PR Newswire copy of the Novo Nordisk OCTANE release. It could not be read from our network (manifest `httpStatus "TypeError"`; first observed in an answer 2026-10-05T14:02:12Z). It is counted as unreadable and never as a qualifying source.

The pull, with redirects resolved, shows when Gemini cited it:

UAE cycle	Admitted?	Gemini answers citing the PR Newswire copy
UAE-C17, 5 Oct	No (the excluded cache cycle, §4.6)	1
UAE-C18, 6 Oct	Yes	1
UAE-C19, 7 Oct	Yes	1
UAE-C20, 8 Oct	Yes	1
UAE-C21, 9 Oct	Yes	1

Under the admission rule, the earliest admitted cycle in which Gemini cited the copy is 6 Oct. Gemini's OCTANE latency from readable sources, 6.2 days, comes from that same cycle. Reading the copy therefore could not shorten Gemini's admitted-cycle OCTANE latency. The 5 Oct citation sits in an excluded cycle and enters no latency.

7.2 E2: naming

7.2.1 Scorer

Naming is decided per answer by the production scorer (`src/engine/score.ts analyzeAnswer().brandMentioned`), unmodified.

The registered measure is brand or molecule. The names counted are the brand, every alias configured on the scan input and, where the scan's domain grants molecule credit, the molecule. As used in every admitted cycle (D24, §11.5.6):

Series	Names counted (registered)	Molecule credit	Brand-only names (D24 sensitivity)
UAE	Wegovy, semaglutide	on (semaglutide)	Wegovy
Singapore	Jardiance, empagliflozin	off (alias list only)	Jardiance
Saudi Arabia	Tzield, teplizumab	off (alias list only)	Tzield

An answer that names only the molecule therefore counts. In the UAE, an answer naming semaglutide in an Ozempic or Rybelsus context also counts. "Claude named Wegovy" means "Claude's answer named Wegovy or semaglutide". The brand-only rate is reported beside the registered rate in §11.5.6. It is a post-registration sensitivity; no bridge offset exists for it.

The scorer's matching internals (for example its tolerance tiers) are withheld under §10. The alias lists above are not internals.

7.2.2 Deterministic consistency check

analyze.ts re-scored every answer of every admitted cycle and compared the per-engine counts with the governed counts stored at scan time: 64 cycles checked, 0 mismatches (scorerValidation). Re-scoring all 64 admitted cycles reproduced the stored naming counts with no mismatch. This verifies deterministic consistency of the pipeline. Naming accuracy was not independently assessed against human-coded answers. The bridge repeated the same check on each source cycle before any paid call (§5.3).

7.2.3 Attribution population

The attribution test splits the engine's after-window governed answers by whether the answer's sha256 is in e1.json qualifyingAnswers. Focused questions are outside the governed population, so qualifying answers to focused questions do not enter the test. Of the 11 qualifying answers:

Series and engine	Qualifying answers	Of which in the governed population
UAE Perplexity	5	1 (the other 4 are focused questions, opaque ids uae-f03 and uae-f04)
UAE Gemini	1	1
Saudi Perplexity	3	3
Saudi Gemini	1	1
Saudi Claude	1	0 (a focused question, opaque id sa-f01)

7.3 E3: investigational-agent intrusion (UAE, semaglutide (Wegovy) series, registered; Singapore, empagliflozin (Jardiance) series, secondary)

Agents and match terms (lib.ts AGENTS): retatrutide (retatrutide, LY-3437943, "triple G"); CagriSema (cagri-sema, cagrilintide, NN-9838); orforglipron (orforglipron, Foundayo, LY-3502970). Every hit is read.

Scheme decisions (verbatim, final coding ledger §2; the ledger's angle-bracket placeholder is shown in square brackets):

- S1.** E3 scope follows protocol §4 exactly: an agent is in scope for a cycle only if it held no approval in the SERIES' market on the cycle date (UAE series → UAE; Singapore series → Singapore). Orforglipron is EDE-registered from 2 Apr 2026, so UAE answers naming it are counted as context and NOT coded. Status in the US or EU stated inside an answer is not scored.
- S2.** Approval status comes only from the regulator snapshot. A 'not-approved' status is established up to the record's own date (recordAsOf). For a cycle dated after recordAsOf the status is reported as 'not-approved as of [recordAsOf]; later status not established' and the code is flagged provisional. This applies to Singapore (HSA listing of 7 Aug 2026) for every cycle.
- S3.** E3 'status correct' requires an explicit statement in the answer that the agent is investigational, in trials, under investigation, not approved, not registered, or not authorised in the market (or generally). 'Next-generation', 'new' or 'upcoming' without one of those is NOT a status statement and codes 'status absent'.
- S4.** A 'which medicine' question is classified from the question text by one fixed rule (lib.ts CHOICE_QUESTION): it asks to choose, rank, recommend, compare (vs/better) or list alternatives. prompt_id is never used to classify, because a prompt_id is a slot.

The pattern of the S4 rule is in lib.ts. It is not reproduced here because its terms paraphrase questions in the withheld battery.

Context. UAE answers naming orforglipron, out of scope by S1: 1,190 (headline.e3.outOfScopeContext). Singapore answers naming an in-scope agent: 0 in every cycle (final coding ledger §4).

Regulator records. Snapshot docs/evidence/computed/easd-2026-regulatory-status-2026-10-11.json, fetched 2026-10-11T16:54:53.745Z, with its previous snapshot of 2026-10-02T10:06:28.749Z. The UAE entries are carried forward from 2 Oct because the EDE directory failed its semaglutide control on 11 Oct (D15, D20). After-window UAE codes are therefore provisional ("not approved as of 2026-10-02; later status not established").

Agent	Market	Status	Approved on	Record as of	Regulator	Carried forward
retatrutide	UAE	not-approved	—	2026-10-02	Emirates Drug Establishment (registered medical product directory)	yes (from 2 Oct)
retatrutide	Singapore	not-approved	—	2026-08-07	Health Sciences Authority (Listing of Registered Therapeutic Products)	no
retatrutide	US	not-approved	—	2026-10-09	FDA (Drugs@FDA via openFDA)	no
retatrutide	EU	not-approved	—	2026-10-11	EMA / European Commission (EMA medicines dataset)	no
cagrisema	UAE	not-approved	—	2026-10-02	Emirates Drug Establishment (registered medical product directory)	yes (from 2 Oct)
cagrisema	Singapore	not-approved	—	2026-08-07	Health Sciences Authority (Listing of Registered Therapeutic Products)	no
cagrisema	US	not-approved	—	2026-10-09	FDA (Drugs@FDA via openFDA)	no
cagrisema	EU	not-approved	—	2026-10-11	EMA / European Commission (EMA medicines dataset)	no
orforglipron	UAE	approved	2026-04-02	2026-10-02	Emirates Drug Establishment (registered medical product directory)	yes (from 2 Oct)
orforglipron	Singapore	not-approved	—	2026-08-07	Health Sciences Authority (Listing of Registered Therapeutic Products)	no
orforglipron	US	approved	2026-04-01	2026-10-09	FDA (Drugs@FDA via openFDA)	no
orforglipron	EU	not-approved	—	2026-10-11	EMA / European Commission (EMA medicines dataset)	no

Ambiguity labels. No separate definition list exists. Each label is defined where it is first used, in the code's note (docs/evidence/easd-2026-manual-codes.json). Labels A2–A10 occur in the frozen record; no label A1 occurs in any frozen file.

Label	Endpoint	Meaning, from the notes	Occurrences in the final codes
A2	E3	"Next-generation drug" or "a trial showed" implies development but states no status; codes <i>status absent</i> (S3)	1 (CagriSema, Claude, 25 Sep)
A3	E3	Listed as an "alternative" with its status stated; stays <i>status correct</i> (D5)	2

Label	Endpoint	Meaning, from the notes	Occurrences in the final codes
A4	E3	Status correct for the agent, but the sentence calls registered products unapproved; outside E3	1
A5	E3	Status stated ("unapproved") but the product class is wrong ("supplement")	1
A6	E5	First-line for high-risk patients with metformin implied or unmentioned; coded high-risk when nothing places it after or on top of metformin; "add-on" with metformin implied codes with-metformin	3 notes
A7	E5	"First-line" <i>within the SGLT2 class</i> , not relative to metformin; codes not-first-line	5 mentions (4 codes and 1 adjudication)
A8	E5	"Monotherapy option" or a label indication is not a first-line position	2
A9	E5	"Preferred choice ... regardless of HbA1c" or "favorable choice" states no position relative to metformin	4 mentions (3 explicit and 1 "as A9 in the baseline")
A10	E3	New in the after window: "research material" or "research compounds ... not established UAE pharmacy treatments", with a warning not to buy; no investigational, trial or unapproved statement, so S3 codes <i>status absent</i>	2 (Perplexity, 8 Oct and 11 Oct)

Strict sensitivity (D16). It removes every code whose note contains the word "ambiguity" (/ambiguity/i). A note that refers to a label without that word (for example "as A9 in the baseline") is not removed. Such notes occur only on E5 not-first-line codes, which the strict count never includes.

7.4 E5: first-line framing (Singapore, empagliflozin (Jardiance) series)

Candidate search (lib.ts E5_CANDIDATE). The protocol's four terms plus equivalents fixed before coding: first-line or first line; initial or initially; instead of, in place of, rather than, ahead of or before metformin; replacing metformin; alternative to metformin; regardless of, independent(ly) of or without metformin; monotherapy. Every hit is read.

Reading unit. Paragraph-level plus headings, with the whole answer opened where unclear (D1, S11).

Codes and order (S10, verbatim):

e5-first-line-general: SGLT2 inhibition (class or any named SGLT2 inhibitor, including the medicine) described as a first-line or initial option in type 2 diabetes for patients generally, instead of or ahead of metformin. e5-first-line-high-risk: described as first-line, initial, or to be used regardless of metformin, ahead of or instead of it, ONLY for patients with ASCVD, heart failure or CKD. e5-first-line-with-metformin: first-line position but together with metformin ('first-line add-on to metformin ... regardless of HbA1c', 'initial combination therapy with metformin'); metformin stays the base, so this is NOT counted as first-line in the protocol sense and is reported separately. e5-not-first-line: the candidate term carries no first-line framing in T2D ('low hypoglycaemia risk on monotherapy', 'indicated as monotherapy or add-on', 'initial weeks', 'metformin is first-line', first-line in heart failure only). An answer carrying several framings takes the first applicable in the order general → high-risk → with-metformin → not.

Protocol count. general + high-risk. With-metformin is reported, never counted (D2). Conventions A6–A9 are in §7.3.

What E5 measures and does not measure. It measures whether an answer frames SGLT2 inhibition as first-line *instead of or ahead of* metformin. It does not measure answers' accuracy about SGLT2 guidance generally, whether an answer recommends the measured medicine, or any prescribing behaviour.

7.5 Saudi arm: E3 and E3-i (post-registration, D10, D12)

E3 agent set (lib.ts T1D_AGENTS; SOURCES in docs/evidence/easd-2026-t1d-agent-set.md):

- SAB-142
- MTX-101
- retogatein (rhGAD65, GAD-alum, "Diamyd", read by hand)
- CRG-002
- CT-868
- hepatic-directed (HDV) insulin lispro
- GZR33 (and the GZR101 premix)
- PRV-101
- CNTY-813
- tipifarnib

Membership rule: "every named molecule or product presented at EASD 2026 in a type 1 or autoimmune diabetes context, clinical or translational, with no FDA or EMA authorisation for any indication". Teplizumab, the measured medicine, is never in the set.

Saudi status is "not established" for every agent: the SFDA searchable register is a POST form behind reCAPTCHA and was not submitted, and the readable SFDA list covers approvals from 1 Feb 2025 to 2 Feb 2026 only. Every agent therefore stays in scope and every code is provisional (T3). Saudi snapshot: docs/evidence/computed/easd-2026-regulatory-status-saudi-t1d-2026-10-05.json, fetched 2026-10-05T13:22:41.937Z, 45 entries:

Agent	Set	Saudi Arabia	US	EU
sab-142	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
mtx-101	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
retogatein	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
crg-002	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
ct-868	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
hdv-insulin-lispro	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
gzh33	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
prv-101	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
cnty-813	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
tipifarnib	E3	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	not-approved; as of 2026-10-02	not-approved; as of 2026-10-05
sotagliflozin	E3-i	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	approved (2023-05-26); as of 2026-10-02	not-approved; as of 2026-10-05
tirzepatide	E3-i	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	approved (2022-05-13); as of 2026-10-02	approved (2022-09-15); as of 2026-10-05
semaglutide	E3-i	a product is on the SFDA new-approvals list (listing date 2026-02-01; not the first registration date); list window to 2026-02-02	approved (2017-12-05); as of 2026-10-02	approved (2018-02-08); as of 2026-10-05

Agent	Set	Saudi Arabia	US	EU
bcg	E3-i	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	approved; as of 2026-10-02	not-established; as of 2026-10-05
fenofibrate	E3-i	not on the SFDA new-approvals list (window 2025-02-01 to 2026-02-02); registration not established	approved (1993-12-31); as of 2026-10-02	approved (2011-04-14); as of 2026-10-05

Reading the Saudi column. The SFDA list the snapshot could read is the *new approvals* list, covering approval dates from 1 Feb 2025 to 2 Feb 2026. It is not the register.

- **Semaglutide.** The date shown (2026-02-01) is the listing date of a generic semaglutide product registered inside that window. It is not the date semaglutide was first registered in Saudi Arabia, which the list cannot show.
- **"Not established."** For tirzepatide and the other molecules this means only that they are absent from that window. It does not mean they are unregistered.
- **Teplizumab.** Its SFDA status comes from the agent-set record (easd-2026-t1d-agent-set.md §4): SFDA news item 18718 of 24 Aug 2025, stage 2 type 1 diabetes, aged 8 and older; the approvals list gives 2025-03-01. The two SFDA dates differ and are not reconciled. SFDA announced the Stage 2 approval (aged 8 and older) in August 2025 (<https://betasfda.sfda.gov.sa/ar/news/18718>); a complete cycle-dated Saudi label was not obtained.

No Saudi status in this study was established from the SFDA register itself (D12).

E3-i, exploratory (T6, verbatim):

molecules authorised for another indication and presented at EASD 2026 as investigational in type 1 diabetes (sotagliflozin, tirzepatide, semaglutide, BCG, fenofibrate). Codes: indication-stated (the answer says the use in type 1 diabetes is off-label, not approved, investigational, or that the molecule is approved only for another condition); indication-absent (named as an option in type 1 diabetes with no such statement); indication-wrong (presented as approved or indicated for type 1 diabetes); not-a-t1d-use (named for its own indication or a non-type-1 population; not scored); not-the-agent.

Sotagliflozin. Zynquista (sotagliflozin) was authorised in the EU on 26 Apr 2019, as an adjunct to insulin in adults with type 1 diabetes. The authorisation was withdrawn on 22 Mar 2022 at the holder's request (European Medicines Agency, Zynquista EPAR; paper reference 94). No EU authorisation is current (agent set §3), so the molecule sits in the E3-i set: it is approved for another indication (heart failure, in the US) and was presented at EASD 2026 as investigational in type 1 diabetes. It was named 3 times before the congress and not after (headline.e3i.byWindowAgent).

Conventions (Saudi manual codes, afterWindowCoder): "a molecule named only to rule it out as a T1D treatment or insulin substitute is not-a-t1d-use; one that also names off-label use is indication-stated".

Sampling (T7, verbatim; the ledger's angle-bracket placeholders are shown in square brackets): "per window and per agent at most 60 naming answers are coded, the lowest ranks of sha256("20261005:[answer sha256]:[agent]"). Unsourced answers are counted as named and shown as NOT-SAMPLED". The hash in the frozen code: const hashOf = (r: SaudiE3iRow) => createHash("sha256").update(`\${E3I\_SEED}:\${r.sha256}:\${r.agent}`).digest("hex"); With const E3I_SEED = "20261005"; (code.ts, lines 883 and 772); answers are ranked by that hex digest, ascending. The rank depends only on the answer, so a window's sample is the same in the baseline and the full run.

Named and sampled per window and agent (headline.e3i.byWindowAgent):

Window	Molecule	Named	Sampled for coding
before-v4	sotagliflozin	3	3
before-v4	tirzepatide	28	28
before-v4	semaglutide	81	60
after	tirzepatide	69	60
after	semaglutide	137	60

Windows and molecules with no naming (BCG, fenofibrate, before-v5, during) are omitted.

Denominator (D16). Sampled namings coded as a type 1 diabetes use (stated, absent or wrong), with distinct answers reported beside it. Namings that only rule the molecule out are excluded. The share is "no status statement" (absent + wrong) / T1D-use namings.

Opportunity rule (D16) for the Saudi E3 zero: answers to questions whose text, with the instruction suffix removed, matches

`\b(new|newest|latest|emerging|upcoming|future|pipeline|research|trials?|investigational|experimental|cure|breakthrough|innovative|advances?|development)\b` (case-insensitive; analyze.ts).

Denominators, kept separate. They are never combined:

Denominator	before-v4	after	Source
Answers searched for E3 namings: every captured answer of the admitted cycles, focused questions included (the E3 denominator)	2446	1789	saudi.opportunity[w].answers (=saudi.e3NamingsByWindow[w].all.answers)
Opportunity answers, within those searched (D16)	44	32	saudi.opportunity[w].opportunityAnswers
Governed E2 population (reference only)	2402	1757	saudi.opportunity[w].governedAnswers
E3-i molecule namings (sampled)	112 (91)	206 (120)	saudi.e3iExploratory[w]
E3-i T1D-use namings (distinct answers)	56 (45)	39 (31)	saudi.e3iExploratory[w]

Non-programme context (T8; counted, never coded): zimislecel 14, baricitinib 4, verapamil 1 answers.

7.6 E4: clinical currency (UAE, semaglutide (Wegovy) series), quarantined

Categories (registered): label · published (not on label) · no identifiable source · congress-window.

Reference values. Fixed in docs/evidence/easd-2026-e4-reference-values.json before any answer was coded (written 2026-10-02T10:08Z; fetched 2 Oct). Sources: the FDA Prescribing Information (NDA 215256 s025), the EMA SmPC section 5.1, PubMed abstracts of the primary publications, and the Novo Nordisk OCTANE release (the only congress-window value; D3). No UAE label could be reached. Scheme decisions S5–S9 (final coding ledger §2) define mean-effect figures, attribution, matching, answer-level precedence and the integer caveat. S7's sentence that OCTANE is not in the readout set is stale; D3 governs (D21 item 6).

D4 hand check. Every figure scripted "no identifiable source" or "congress-window", in both windows, was read before publication.

- First read: a separate agent of the same model family.
- Adjudication: the study coordinator adjudicated every uncertain item, all six congress-window items, and random samples of 11 exclusions and 30 no-source confirmations. One confirmation was overturned.
- The 12 figures from the 11 Oct UAE cycle were read by the coordinator directly (manual-codes.json e4Checker).
- At the D17 entry (through 10 Oct): 195 figures read, 104 excluded.
- Final record through 11 Oct (headline.e4.d4; stated in D21 item 5): 113 of 207 hand-checked figures excluded on reading, 0 recategorised, 0 unchecked. Retained: 92 no-source figures and 2 congress-window figures.
- Of the six congress-window matches, four were Rybelsus comparators and two are genuine OCTANE (D17).

D4 check records by reason code (docs/evidence/easd-2026-manual-codes.json e4; 204 check records, because a figure repeated verbatim in one answer carries one check: 93 confirmed, 111 excluded):

Reason code	Excluded check records
P2	33
P3	32
M2	20
P1	18
S1	4
M1	3
O1	1

Confirmed check records: 93.

D17 audit of the label and published attributions. Rule, fixed before the sample was drawn: a seeded random sample (seed 20261010) of 150 of the 1,082 figures matched to a label or published value, read under the same rules, with a 5% bar.

Reading	Population	Sample	Wrongly attributed	Share	Wilson 95%
First read (separate agent, same model family)	1082	150	17	0.1133	0.072–0.174
Adjudicated (study coordinator)	1082	150	12	0.08	0.0464–0.1346

Wrongly categorised: none. First-read verdicts overruled at adjudication: 5 (wrong-attribution → correct; "typical range" figures that S7 codes endpoint by endpoint, D17). Adjudicated wrong attributions by reason code: P3 6, P1 3, M2 2, M1 1. Verdict, verbatim: "FAILS the 5% bar under either reading: the label/published split of E4 is not published as coded (D17)."

Why E4 is quarantined. Under either reading the error rate exceeds the 5% bar. By D17's rule, the label/published split is not published as coded. E4 reports only the hand-verified counts, the audited error rate with its interval, and the statement that the scripted categories failed audit. No efficacy-source comparison is published.

8. Statistics

8.1 E2 decision

A movement is declared when the after-window pooled governed rate lies outside the bridged band (I1). The decision is deterministic given the band and the offset bounds; it carries no p-value.

8.2 Attribution test (I3)

Fisher's exact test, two-sided, on the 2 × 2 table:

	Named	Not named
E1-qualifying answers	a	b
Other answers	c	d

Both groups are the engine's after-window governed answers.

- Implementation: `analyze.ts fisher`. The two-sided p sums the hypergeometric probabilities of every table, with the same margins, that is no more probable than the observed table (tolerance factor $1 + 10^{-9}$).
- The test is not run when either group is empty (p shown as —).
- A *congress effect* (reported as "protocol-defined congress-linked movement") requires a declared movement, $E1 > 0$, and $p < 0.05$.
- It is evidence of an observable connection to congress sources, not a demonstration that the congress produced the movement; model updates and other sources can move the same answers (D16).
- **Minimum attainable p (D24, §11.5.8; frozen E1 classification).** With few qualifying answers, no split could reach $p < 0.05$. This is the smallest two-sided p any split of the qualifying answers could give with both margins fixed:
 - UAE Perplexity: 1 qualifying answer, minimum p 0.3571.
 - UAE Gemini: 1 qualifying answer, minimum p 0.455. Under the corrected classification (§10.2): 2 qualifying answers, both named, against 204 of 376 others; minimum p 0.206, Fisher p 0.503.
 - Saudi Gemini: 1 qualifying answer, minimum p 0.0909.
 - Saudi Perplexity: 3 qualifying answers, minimum p 0.0002, observed p 0.1674. This is the only case where the test could have been informative.

Where an engine had no qualifying answer, its reading rests on $E1 = 0$ alone. The test also excludes focused questions, which removes 4 of UAE Perplexity's 5 qualifying answers and Saudi Claude's only one (§7.2.3). Removing answers that cite unverified pages from the comparison group changes no reading.

8.3 Wilson intervals

Wilson score interval with $z = 1.959964$ (`code.ts wilson`), clipped to $[0, 1]$. Used for the D17 audit, and for the Saudi ledger's pooled E3-i share. That pooled share treats namings as independent although several share an answer, and it combines both windows. It is shown in the ledger only and is not an analysis figure.

Single-cycle Wilson intervals (the per-cycle view, §11.5.4) are illustrative binomial intervals. Their interpretation assumes an appropriate independent Bernoulli sampling model, which this fixed prompt battery does not establish. Repeated-question dependence additionally affects pooled and longitudinal comparisons. These intervals are not used to validate the registered movement decision.

8.4 Question-clustered bootstrap intervals (robustness, D16)

For each after-window rate, answers are grouped by question (prompt id). Questions are resampled with replacement, 2,000 times. The pooled share is K/N over the resampled questions. The interval is the 2.5th and 97.5th percentile order statistics: lower at $\text{index floor}(0.025 \times 2000) = 50$, upper at $\text{index ceil}(0.975 \times 2000) - 1 = 1949$, zero-based.

PRNG: mulberry32. Seeds (`analyze.ts`):

Analysis	Seed rule	Seeds
E2, UAE	20261012 + engine index (Perplexity 0, Gemini 1, ChatGPT 2, Claude 3; the four-engine row -1)	Perplexity 20261012, Gemini 20261013, ChatGPT 20261014, Claude 20261015, four engines 20261011
E2, Saudi	the same + 10	Perplexity 20261022 ... four engines 20261021
Saudi E3-i	20261012 + 99 + window index	before-v4 20261111, after 20261114

The bridge bootstrap is described in §5.2.

8.5 Cohen's kappa (second pass)

Unweighted Cohen's kappa between the first code and the blind second-pass code, over the categories used by either coder in that endpoint, computed before adjudication. Agreement is the share of identical codes. No script in `scripts/easd-2026-study/` computes these statistics; they are stored in `docs/evidence/easd-2026-second-pass.json` stats. For this supplement they were recomputed from that file's `items`, and all three kappas and agreements reproduce: E5 0.9768 and 0.9894; E3 1 and 1; E3-i 0.8663 and 0.916.

8.6 The house magnitude scale

Label	Change
SMALL	< 5pp
POTENTIALLY RELEVANT	5 to < 10pp
DECISION-RELEVANT	10 to < 15pp
MAJOR	≥ 15pp

Applied to the difference between the after-window pooled rate and the bridged pre-congress reference (I2). The scale is a description of size. It is not a significance test, not the movement rule (the band is), and not an action threshold.

8.7 Detectable effect (registered)

95% half-width at $p = 0.5$: $\approx \pm 9.4\text{pp}$ for about 108 answers per engine; $\approx \pm 5.0\text{pp}$ for about 378. The UAE after window delivered 378 governed answers per engine (seven cycles $\times$ 54).

9. Validation, human oversight and review

9.1 Deterministic consistency checks

- Analysis: 64 admitted cycles re-scored, 0 mismatches against the governed counts (`analysis.json` `scorerValidation`).
- Bridge: each source cycle re-scored and matched before any paid call (`bridge.json` `validation.matchesGovernedCounts: true` for all three series).
- Coding consistency: `verify.ts` checks the coder's JSON against its ledger (admissions, per-cycle tables, sums, no UNCODED). Its full output for the regenerated final outputs is frozen at `docs/evidence/computed/easd-2026-verify.txt` (D21 item 7): 38 checks PASS, 0 FAIL, and 1 informational line ("after-window cycles admitted: UAE 7, Singapore 7"). Final line: "all checks passed". The baseline lines count 15 admitted UAE cycles through 29 Sep (13 before-v4, 1 before-v5, 1 during).

9.2 The blind second coding pass (robustness, D16)

Source: `docs/evidence/easd-2026-second-pass.json`.

Selection and procedure.

Blind second coding pass (prereg D16 robustness): a separate instance of the same model family coded every E3 code, every non-trivial E5 code plus a seeded sample of 20 not-first-line codes, and every sampled E3-i T1D-use or not-the-agent code plus a seeded sample of 20 not-a-t1d-use codes, through 10 Oct, from the text alone under the written rules, without the first codes. It is not independent external validation. Top-up of 11 Oct, same rules and same blind instructions: every new E3 code, every new non-trivial E5 code (none on 11 Oct), every newly sampled E3-i T1D-use code, and a seeded sample of 5 not-a-t1d-use codes (seed 20261011). Adding the 11 Oct namings displaced 14 earlier blind-coded E3-i items from the capped sample; they stay in this record and in the agreement statistics, which measure coder agreement, not the sample.

The file records one seed field, 20261011; D21 item 6 states that both batches used it. Both worksheet builders were run as inline scripts. Their text is committed in `docs/evidence/easd-2026-second-pass-builders.py` and is reproduced verbatim, apart from the function wrappers and two omitted diagnostic prints (the top-up's pool sizes and its count of 14 displaced items, both recorded in `easd-2026-second-pass.json`). The batch-1 builder calls `random.seed(20261011)` before drawing the E5 sample of not-first-line codes, then draws the E3-i sample and shuffles the items. The builders read local temp files (the pull and the coder outputs of the time), which are not published. The blind coder received only the worksheet and the written rules; the key stayed with the coordinator.

Batches.

Batch	E3 items (agree)	E5 items (agree)	E3-i items (agree)	Total (agree)
Through 10 Oct	18 (18)	94 (93)	117 (108)	229 (219)
Top-up 11 Oct	3 (3)	0	14 (12)	17 (15)
All	21 (21)	94 (93)	131 (120)	246 (234)

Agreement (stats):

Endpoint	Items	Agreement	Cohen's kappa	Disagreements
E5	94	0.9894	0.9768	1
E3	21	1	1	0
E3-i	131	0.916	0.8663	11

Cross-tabulation of first versus second code (from items):

Endpoint	First code → second-pass code	Items
E3	status-correct → status-correct	18
E3	status-absent → status-absent	3
E5	e5-not-first-line → e5-not-first-line	28
E5	e5-first-line-with-metformin → e5-first-line-with-metformin	63
E5	e5-first-line-high-risk → e5-first-line-high-risk	2
E5	e5-first-line-general → e5-not-first-line	1
E3-i	indication-stated → indication-stated	36
E3-i	indication-absent → indication-absent	61
E3-i	not-a-t1d-use → not-a-t1d-use	22
E3-i	indication-absent → indication-stated	2
E3-i	indication-absent → not-a-t1d-use	6
E3-i	not-a-t1d-use → indication-stated	3
E3-i	not-the-agent → not-the-agent	1

Every disagreement and its adjudication. The study coordinator adopted the second-pass code 5 times and kept the first code 7 times (adoptedFromSecondPass, keptFirstCode).

id	Endpoint	Batch	Agent	First code	Second-pass code	Second-pass note	Adjudicated	Adjudication
30	E3-i	through 10 Oct	tirzepatide	indication-absent	indication-stated	uncertain: GLP-1 "may be considered only in specialised circumstances ... they are primarily type 2 diabetes or obesity treatments"	indication-stated	Adopted the second-pass code: "they are primarily type 2 diabetes or obesity treatments" states the indication, so the T1D use is framed as outside it.
35	E3-i	through 10 Oct	tirzepatide	indication-absent	not-a-t1d-use	uncertain: "Out-of-pocket burden deters uptake even when medications like Tirzepatide show efficacy" is a cost/access remark, no T1D use offered	not-a-t1d-use	Adopted the second-pass code: a cost and access remark; no T1D use is offered.
50	E3-i	through 10 Oct	semaglutide	indication-absent	not-a-t1d-use	"Hikma launched Glucentra (semaglutide), the first generic version of Ozempic" is a market launch	indication-absent	Kept: the molecule sits under "GLP-1 Receptor Agonists (Newer Option)" in a list of type 1 treatment options, so it is offered as an option.
66	E3-i	through 10 Oct	semaglutide	indication-absent	not-a-t1d-use	"Semaglutide was the largest segment with 73.9% revenue share" is a market statistic	indication-absent	Kept: listed as a heading entry in a type 1 options answer.
74	E3-i	through 10 Oct	semaglutide	not-a-t1d-use	indication-stated	"Not indicated for use in type 1 diabetes mellitus"	not-a-t1d-use	Kept: a pure rule-out quoting the label limitation; rule-outs without off-label naming are not-a-t1d-use in both windows.
102	E3-i	through 10 Oct	semaglutide	not-a-t1d-use	indication-stated	uncertain: [Gemini text not reproduced, §13.3]	not-a-t1d-use	Kept: named for type 2 and the landscape, not offered for type 1.
113	E3-i	through 10 Oct	semaglutide	indication-absent	indication-stated	uncertain: GLP-1 "may be considered only in specialised circumstances ... primarily type 2 diabetes or obesity treatments"	indication-stated	Adopted the second-pass code: "they are primarily type 2 diabetes or obesity treatments" states the indication, so the T1D use is framed as outside it.
119	E3-i	through 10 Oct	tirzepatide	indication-absent	not-a-t1d-use	The second coder interpreted the passage as a market-availability remark.	indication-absent	Kept: the adjudicator interpreted the molecule's placement as a heading in a type 1 options answer, offering a type 1 option.
140	E3-i	through 10 Oct	tirzepatide	indication-absent	not-a-t1d-use	The second coder read the passage as a rule-out or specialist caution only.	indication-absent	Kept: the adjudicator treated the passage as conditionally offering adjunct use in type 1 diabetes, as in items the second pass itself coded absent.
222	E5	through 10 Oct	—	e5-first-line-general	e5-not-first-line	uncertain: "Jardiance is often chosen as a first-line treatment" compared with Invokana, read as a within-class brand ranking	e5-not-first-line	Adopted the second-pass code: "first-line" here ranks Jardiance against Invokana within the SGLT2 class, which the baseline convention A7 codes not-first-line in its two other cases; applied consistently.
232	E3-i	top-up 2026-10-11	semaglutide	not-a-t1d-use	indication-stated	The second coder read a phrase placing the approved use in type 2 diabetes as stating that the type 1 use lies outside the indication.	not-a-t1d-use	Kept the first code: the sentence is a market remark on adoption and prescription growth that offers no type 1 use, so the indication cue does not make it indication-stated (as 19fc3dbc, same wording).
234	E3-i	top-up 2026-10-11	semaglutide	indication-absent	not-a-t1d-use	'do not replace insulin in type 1 diabetes' and 'specialist-directed and may be unsuitable' only rule the drugs out.	not-a-t1d-use	Adopted the second-pass code: "Their use in T1D is specialist-directed and may be unsuitable" is the rules' rule-out example ("should not be used without a specialist"), not an offered T1D use.

Displaced items. The 14 earlier blind-coded E3-i items displaced from the capped sample by the 11 Oct namings remain in the agreement statistics but not in the analysis sample. The statistics therefore describe coder agreement, not the coded sample.

What the pass measures. Consistency between two instances of the same model family reading the same text under the same written rules. It is not independent clinical validation, and it is not a human reading.

9.3 The D4 hand check

See §7.6.

9.4 The E4 audit

See §7.6. Reason codes (e4ReasonCodes, also used in the D4 records):

Code	Meaning
P1	different product, molecule or intervention
P2	comparator effect attributed to the target product
P3	joint or class-level figure without valid brand attribution
M1	responder proportion mistaken for a mean effect
M2	maximum, subgroup or illustrative result mistaken for a mean effect
S1	incorrect congress-source association
O1	other S5 exclusion (BMI value, price)
U1	insufficient evidence to verify attribution

U1 was not assigned in either the D4 records or the audit.

9.5 Research Governance, Author Contributions and AI Assistance

This research was conducted under the direction of the author through AlphaCitation's research programme. AI-assisted tools supported data processing, coding, analytical implementation, verification and manuscript preparation. The study documents its research protocol, methodological amendments, validation procedures and reproducibility controls. The named author assumes responsibility for the final interpretation and publication.

The specifics, from the frozen record:

Step	Who or what performed it, as recorded
Protocol, deviations, scripts	Drafted with Anthropic Claude models (commits co-authored "Claude Opus 5.5"), under the author's direction
Baseline E3 and E5 codes (2–3 Oct)	"signal-note agent (Claude), read in full" (manual1-codes.json coder)
After-window E3 and E5 codes; Saudi E3-i codes; E1 hand ledger; adjudications of the second pass, D4 and D17	"Study coordinator (Claude Opus 5.5)"
D4 first read; D17 first read; blind second pass	Separate agents of the same model family
E2 naming	The production scorer (deterministic software); its re-scoring was checked for deterministic consistency against the governed counts
Founder decisions recorded in the protocol	Approval of the daily cadence on 5 Oct (D9); the Saudi arm at the founder's request (D10); §6 required founder approval of spend, and the bridge ran, but that approval is not dated in the frozen record; naming the three measured medicines in the publication (D22); publication and review policy (§9.7)

Funding and Competing Interests

This study was independently conducted and funded by AlphaCitation. No pharmaceutical manufacturer commissioned, sponsored, or influenced the research.

The author is the founder of AlphaCitation, which develops and offers AI information intelligence services to life-sciences organizations. He previously held commercial leadership positions at Novo Nordisk, whose medicine Wegovy is included in the study.

Naming the measured medicines (D22). The publication names semaglutide (Wegovy, UAE), empagliflozin (Jardiance, Singapore) and teplizumab (Tzield, Saudi Arabia), consistently and whatever their results.

- The decision was made on 11 Oct, after preliminary and final results were available.
- It is not part of the registered protocol: the registered §10 had withheld the medicines.
- It is not neutral in perception. Naming a medicine implies no endorsement, sponsorship or commercially meaningful effect.
- It changes no endpoint, result or rule (§2.3 gives the full history).

The subject of every finding is what the engines' answers contained: naming, citation, status statements, framing, and their fidelity to the evidence and to the regulator records. No finding is about the medicines themselves.

Human oversight was the author's. The frozen codes, ledger entries and adjudications were model-assisted, as listed. After the freeze, the author checked a sample of 50 interpretive codes blind (§9.6); no frozen code was changed. The blind second coder is the same model family as the first, so the second pass measures consistency, not independent validation.

9.6 Human validation of the interpretive coding

The study had **blinded human validation of the interpretive coding (50 sampled E3, E5 and E3-i judgements), by the author as a single reader.**

- **It is not independent or external validation.** The reader is the study's author. The reader coded blind to the original codes, the answering engine and the date, but is not independent of the study.
- **It is a targeted quality-assurance check,** not a research endpoint and not a clinical review.
- **The naming measure (Part A) was not human-coded.** Its re-scoring (64 admitted cycles, 0 mismatches, §9.1) is a deterministic consistency check; naming accuracy was not independently assessed against human-coded answers. The per-answer outcome table reproduces every governed per-cycle count (§11.5.10).
- **No frozen code or registered figure was changed by the check.** A disagreement is reported as a finding.

What the agreement figures mean. In selected blinded quality checks, the author agreed with 17/17 codes in the ordinary sampled strata and 30/33 codes in deliberately risk-enriched strata. These are quality-check results for the selected items, not estimates of overall coding accuracy or independent clinical validation. No pooled agreement figure or interval is reported for the 50 items.

Method. The protocol is docs/research/easd-2026/human-validation-protocol.md, fixed on 11 Oct before any item was coded.

- The sample was drawn by fixed rules with seed 20261012. The representative sample and the risk-enriched sample (hard cases, expected to agree less) are kept distinct and never pooled.
- Items were shuffled and given opaque IDs (A001–A100, B001–B050).
- For each item the reader saw the market, the question with its market-framing instruction removed, the reading unit (the answer's headings and every paragraph carrying the term, with the full answer available), the code options with their written definitions, and a note box.

- Withheld from the reader: the original code and its note, the engine and model, the cycle, date and capture time, the answer hash, the prompt slot, the stratum and the selection rules.
- Scoring compared the reader's code with the frozen final code.

Part B strata, from the protocol:

Endpoint	Stratum	Sample	Rule	Items
E3 (UAE)	Clinically material	risk-enriched	census of status-absent or status-wrong in which-medicine questions (supplement §7.3, S4)	3
E3 (UAE)	Status-correct sample	representative	seeded sample of status-correct (of 18)	5
E5 (Singapore)	Protocol-counted	risk-enriched	census of e5-first-line-general and e5-first-line-high-risk (0 and 2 in the frozen codes)	2
E5 (Singapore)	Ambiguity note	risk-enriched	census of every other code whose note records an ambiguity or a convention label (/ambiguity bA[6-9]b/): 10 e5-not-first-line, 1 e5-first-line-with-metformin	11
E5 (Singapore)	With-metformin sample	representative	seeded sample of the remaining e5-first-line-with-metformin codes	6
E3-i (Saudi Arabia)	Rule-out plus off-label	risk-enriched	census of sampled codes whose reading unit both rules the molecule out as an insulin substitute and names off-label use (4 indication-stated, 2 not-a-t1d-use)	6
E3-i (Saudi Arabia)	Specialist use	risk-enriched	seeded, one per code in turn (indication-stated, indication-absent, not-a-t1d-use): reading unit mentions a specialist, an endocrinologist or "selected cases / circumstances / adults / patients"	6
E3-i (Saudi Arabia)	Past-tense research	risk-enriched	seeded, one per code in turn: "studied", "has / have been investigated / evaluated / explored / studied", "studies ... explored / evaluated / investigated", "evaluated / explored the use"	5
E3-i (Saudi Arabia)	Type-1-use sample	representative	seeded, one per code in turn (indication-stated, indication-absent), no borderline flag	4
E3-i (Saudi Arabia)	Not-a-type-1-use sample	representative	seeded not-a-t1d-use, no borderline flag	2

Scope, by the author's decision. Part B, the interpretive judgements, was coded in full: 50 of 50 items. Part A, the 100 naming observations, was not coded.

Provenance (docs/research/easd-2026/human-validation-results.md, generated by the scoring script from the reader's exported codes and the private key; item IDs, codes and the reader's notes only, no question or answer text):

- Reader's CSV sha256: 4b16d70202fe47d4c218669736d484c96420d03ebc44b4890f666dfd0f44b4cb
- Key seed: 20261012; worksheet sha256 recorded in the key: 2750d9b69ba10e976bdc78464c84d8975e86b377e6f875556c1ba9cf776a5be1 (matches the key)
- Frozen pull sha256: de98a60469b0e7b100110f16170b16760cc32fe333696eb4292bf48219eed265
- Scope decided by the author: Part B (interpretive judgements) only. Part A (naming) was not coded; the naming scorer is deterministic and reproduces every governed count.
- Items coded: Part A 0 of 100; Part B 50 of 50. The 100 Part A items (A001-A100) were not coded and are excluded from every figure.

Results, Part B (frozen final code against the reader):

Endpoint and sample	Items	Agreements	Agreement	Wilson 95%	Cohen's kappa
All endpoints, representative	17	17	100.0%	81.6%–100.0%	1.000
All endpoints, risk-enriched	33	30	90.9%	76.4%–96.9%	0.886
E3, representative	5	5	100.0%	56.6%–100.0%	n/a (one category only)
E3, risk-enriched	3	3	100.0%	43.9%–100.0%	n/a (one category only)
E5, representative	6	6	100.0%	61.0%–100.0%	n/a (one category only)
E5, risk-enriched	13	13	100.0%	77.2%–100.0%	1.000
E3-i, representative	6	6	100.0%	61.0%–100.0%	1.000
E3-i, risk-enriched	17	14	82.4%	59.0%–93.8%	0.718

Kappa is computed within each row over the codes either party used. Across endpoints the code sets are disjoint, so the all-endpoint kappa is a summary only. Kappa is undefined or low when one code dominates a row, even when every judgement agrees.

Cross-tabulations, both samples (rows: reader; columns: original code):

Reader ↓ / original →	status-correct	status-absent
status-correct	5	0
status-absent	0	3

Reader ↓ / original →	e5-first-line-high-risk	e5-first-line-with-metformin	e5-not-first-line
e5-first-line-high-risk	2	0	0
e5-first-line-with-metformin	0	7	0
e5-not-first-line	0	0	10

Reader ↓ / original →	indication-stated	indication-absent	indication-wrong	not-a-t1d-use
indication-stated	10	0	0	2
indication-absent	0	6	0	0
indication-wrong	0	1	0	0
not-a-t1d-use	0	0	0	4

Every disagreement. All three are risk-enriched, borderline Saudi E3-i items (an exploratory endpoint). Two of them are the same answer, read once per molecule. No frozen code was changed. The reader left no note on any of them. No answer text is reproduced.

Item	Sample	Judgement	Stratum	Selection rules met	Market	Engine	Window	Reader's code	Original (frozen) code	Answer sha256 (first 12)	Reader's note
B014	risk-enriched	E3-i (semaglutide)	B-E3i-specialist	specialist	Saudi Arabia	Gemini	after	indication-wrong	indication-absent	c5a9dda0d305	—
B035	risk-enriched	E3-i (semaglutide)	B-E3i-ruleout-offlabel	ruleout-offlabel, specialist	Saudi Arabia	Perplexity	after	indication-stated	not-a-t1d-use	ca6baa411be0	—
B039	risk-enriched	E3-i (tirzepatide)	B-E3i-ruleout-offlabel	ruleout-offlabel, specialist	Saudi Arabia	Perplexity	after	indication-stated	not-a-t1d-use	ca6baa411be0	—

9.7 External review

Version 1.0 is published without mandatory external pre-publication review, after passing AlphaCitation's internal methodological, analytical, factual and editorial quality gates. It invites scholarly and practitioner feedback after publication, and substantive criticism, corrections and revised editions are recorded in a transparent version history.

- No independent external peer review, journal review, external clinical review or external statistical review of this study, its analysis or this supplement is recorded as of v1.0.
- External critical review of the release candidate.** Before release, an external critical review of the candidate paper, supplement and brief was commissioned. It was not a peer review and is not published. Its corrections were applied without changing any frozen value; they include the corrected E1 either-route tally shown beside the frozen one (§10.2), narrower claim wording on attribution, detection timing, validation and inspection coverage, the title, and the disclosure, licensing and availability statements.
- The frozen record notes one external check of a single E1 source: an external reviewer's independent check that found the unnamed TRIUMPH-2 item in the diaTribe round-up (D18). Its outcome is recorded in §7.1.7. It was not a review of the study.
- Internal quality checks (not external review).** The publication passed an internal claim-to-evidence check (251 claims in its second round) and three internal perspective reviews (methodological, pharmaceutical, GEO); they are not published. Their outcomes:
 - First round of the claim check: 214 claims, 186 verified; 3 critical, 8 major and 17 minor defects.
 - Second round: 251 claims, 232 verified; 1 critical, 5 major and 13 minor defects. Of the 28 first-round defects, 25 were resolved and 3 partly resolved.
 - Methodological review: major revision.
 - Pharmaceutical review: serious and honest, but not shareable as written.
 - GEO review: a careful public treatment of its question, with five corrections.
 - The resulting changes are recorded as D24 (§2.2, §11.5) and in this supplement.
- Any external review that takes place will be recorded, with its scope and outcome, in a later version of this supplement.

AlphaCitation Research is committed to open scientific scrutiny. We welcome methodological critiques, requests for clarification and substantive commentary on our published research. Material corrections and updates will be documented in any subsequent versions.

Correspondence.

Scholarly correspondence, methodological critiques and correction requests may be directed to analyst@alphacitation.com. Substantive corrections will be evaluated and documented in the publication's version history.

10. Results tables in full

Each table names its file and field. Tables marked "(verbatim)" are copied from the named Markdown output unchanged.

10.1 E1: answers citing a congress-derived source or reporting a readout result, per cycle (verbatim)

Source: docs/evidence/easd-2026-e1.md; fields perEngineCycle[series][slug][engine].either, .n and .unverified in docs/evidence/computed/easd-2026-e1.json. "+N unverified": answers that cite at least one unreadable new URL and do not qualify.

uae-gl1p1

Cycle	Date	Window	Perplexity	Gemini	ChatGPT	Claude
UAE-C01	2026-08-22	before-v4	0/54	0/60	0/60	0/60
UAE-C02	2026-08-24	before-v4	0/60	0/60	0/60	0/60
UAE-C03	2026-08-26	before-v4	0/60	0/60	0/60	0/60
UAE-C04	2026-08-28	before-v4	0/60	0/60	0/60	0/60
UAE-C05	2026-08-30	before-v4	0/59	0/60	0/60	0/60
UAE-C06	2026-09-04	before-v4	0/60	0/60	0/60	0/60
UAE-C07	2026-09-09	before-v4	0/60	0/60	0/60	0/60
UAE-C08	2026-09-11	before-v4	0/60	0/60	0/60	0/60
UAE-C09	2026-09-14	before-v4	0/57	0/59	0/60	0/60
UAE-C10	2026-09-16	before-v4	0/56	0/60	0/60	0/60
UAE-C11	2026-09-21	before-v4	0/59	0/60	0/60	0/60
UAE-C12	2026-09-23	before-v4	0/60	0/60	0/60	0/60
UAE-C13	2026-09-25	before-v4	0/60	0/60	0/60	0/60
UAE-C14	2026-09-27	before-v5	0/58	0/60	0/60	0/60
UAE-C15	2026-09-29	during	0/60 (+8 unverified)	0/60 (+44 unverified)	0/60 (+5 unverified)	0/60 (+6 unverified)
UAE-C16	2026-10-04	after	3/60 (+8 unverified)	0/60 (+44 unverified)	0/60 (+4 unverified)	0/60 (+14 unverified)
UAE-C18	2026-10-06	after	2/60 (+9 unverified)	1/60 (+45 unverified)	0/60 (+7 unverified)	0/60 (+9 unverified)
UAE-C19	2026-10-07	after	0/60 (+7 unverified)	0/60 (+45 unverified)	0/60 (+4 unverified)	0/60 (+10 unverified)
UAE-C20	2026-10-08	after	0/60 (+5 unverified)	0/60 (+43 unverified)	0/60 (+8 unverified)	0/60 (+11 unverified)

Cycle	Date	Window	Perplexity	Gemini	ChatGPT	Claude
UAE - C21	2026-10-09	after	0/60 (+7 unverified)	0/60 (+46 unverified)	0/60 (+6 unverified)	0/60 (+10 unverified)
UAE - C22	2026-10-10	after	0/60 (+29 unverified)	0/60 (+45 unverified)	0/60 (+9 unverified)	0/60 (+9 unverified)
UAE - C23	2026-10-11	after	0/60 (+27 unverified)	0/60 (+51 unverified)	0/60 (+16 unverified)	0/60 (+13 unverified)

singapore-sg1t2

Cycle	Date	Window	Perplexity	Gemini	ChatGPT	Claude
SG-C01	2026-08-22	before-v4	0/22	0/26	0/26	0/26
SG-C02	2026-08-24	before-v4	0/26	0/26	0/26	0/26
SG-C03	2026-08-26	before-v4	0/26	0/26	0/26	0/26
SG-C04	2026-08-28	before-v4	0/26	0/26	0/26	0/26
SG-C05	2026-08-30	before-v4	0/26	0/26	0/26	0/26
SG-C06	2026-09-04	before-v4	0/26	0/26	0/26	0/26
SG-C07	2026-09-06	before-v4	0/26	0/26	0/26	0/26
SG-C08	2026-09-08	before-v4	0/26	0/26	0/26	0/26
SG-C09	2026-09-10	before-v4	0/26	0/26	0/26	0/26
SG-C10	2026-09-13	before-v4	0/26	0/26	0/26	0/26
SG-C11	2026-09-15	before-v4	0/26	0/26	0/26	0/26
SG-C12	2026-09-17	before-v4	0/26	0/26	0/26	0/26
SG-C13	2026-09-19	before-v4	0/26	0/25	0/26	0/26
SG-C14	2026-09-21	before-v4	0/25	0/26	0/26	0/26
SG-C15	2026-09-26	before-v4	0/26	0/26	0/26	0/26
SG-C17	2026-10-02	during	0/26 (+7 unverified)	0/26 (+15 unverified)	0/26 (+1 unverified)	0/26
SG-C18	2026-10-05	after	0/26 (+6 unverified)	0/26 (+18 unverified)	0/26 (+1 unverified)	0/26
SG-C19	2026-10-06	after	0/26 (+6 unverified)	0/26 (+10 unverified)	0/26 (+4 unverified)	0/26 (+1 unverified)
SG-C20	2026-10-07	after	0/26 (+3 unverified)	0/26 (+17 unverified)	0/26 (+3 unverified)	0/26 (+1 unverified)
SG-C21	2026-10-08	after	0/26 (+3 unverified)	0/26 (+14 unverified)	0/26 (+4 unverified)	0/26
SG-C22	2026-10-09	after	0/26 (+2 unverified)	0/26 (+13 unverified)	0/26 (+1 unverified)	0/26 (+1 unverified)
SG-C23	2026-10-10	after	0/26 (+3 unverified)	0/26 (+16 unverified)	0/26 (+4 unverified)	0/26 (+2 unverified)
SG-C24	2026-10-11	after	0/26 (+4 unverified)	0/26 (+19 unverified)	0/26 (+4 unverified)	0/26 (+2 unverified)

saudi-t1d (post-registration, D10)

Cycle	Date	Window	Perplexity	Gemini	ChatGPT	Claude
SA-C01	2026-08-24	before-v4	0/56	0/56	0/56	0/56
SA-C02	2026-08-27	before-v4	0/56	0/56	0/56	0/56
SA-C03	2026-08-30	before-v4	0/54	0/55	0/56	0/56
SA-C04	2026-09-01	before-v4	0/56	0/56	0/56	0/56
SA-C05	2026-09-03	before-v4	0/56	0/56	0/56	0/56
SA-C06	2026-09-08	before-v4	0/55	0/56	0/56	0/56
SA-C07	2026-09-10	before-v4	0/56	0/56	0/56	0/56
SA-C08	2026-09-14	before-v4	0/55	0/56	0/56	0/56
SA-C09	2026-09-16	before-v4	0/53	0/55	0/56	0/56
SA-C10	2026-09-21	before-v4	0/51	0/56	0/56	0/56
SA-C11	2026-09-26	before-v4	0/52	0/56	0/56	0/56
SA-C13	2026-10-04	after	0/56 (+10 unverified)	0/56 (+52 unverified)	0/56 (+3 unverified)	0/56 (+10 unverified)
SA-C14	2026-10-05	after	2/56 (+12 unverified)	0/56 (+50 unverified)	0/56 (+5 unverified)	0/53 (+7 unverified)
SA-C15	2026-10-06	after	1/56 (+16 unverified)	0/56 (+52 unverified)	0/56 (+3 unverified)	0/56 (+6 unverified)
SA-C16	2026-10-07	after	0/56 (+12 unverified)	1/56 (+53 unverified)	0/56 (+2 unverified)	0/56 (+8 unverified)
SA-C17	2026-10-08	after	0/56 (+5 unverified)	0/56 (+50 unverified)	0/56 (+3 unverified)	1/56 (+7 unverified)
SA-C18	2026-10-09	after	0/56 (+3 unverified)	0/56 (+50 unverified)	0/56 (+5 unverified)	0/56 (+12 unverified)
SA-C19	2026-10-10	after	0/56 (+3 unverified)	0/56 (+54 unverified)	0/56 (+5 unverified)	0/56 (+11 unverified)
SA-C20	2026-10-11	after	0/56 (+4 unverified)	0/56 (+50 unverified)	0/56 (+6 unverified)	0/56 (+11 unverified)

10.2 E1 by window and engine (sums of the per-cycle cells)

Source: sums over perEngineCycle by windows[series|slug].window. All before-window cells are 0.

Series	Window	Engine	Answers	Qualifying (either route)	Via source	Via text	Unverified
uae-glp1	during	Perplexity	60	0	0	0	8
uae-glp1	during	Gemini	60	0	0	0	44
uae-glp1	during	ChatGPT	60	0	0	0	5
uae-glp1	during	Claude	60	0	0	0	6
uae-glp1	after	Perplexity	420	5	5	0	92
uae-glp1	after	Gemini	420	1	1	0	319
uae-glp1	after	ChatGPT	420	0	0	0	54
uae-glp1	after	Claude	420	0	0	0	76

Series	Window	Engine	Answers	Qualifying (either route)	Via source	Via text	Unverified
singapore-sglt2	during	Perplexity	26	0	0	0	7
singapore-sglt2	during	Gemini	26	0	0	0	15
singapore-sglt2	during	ChatGPT	26	0	0	0	1
singapore-sglt2	during	Claude	26	0	0	0	0
singapore-sglt2	after	Perplexity	182	0	0	0	27
singapore-sglt2	after	Gemini	182	0	0	0	107
singapore-sglt2	after	ChatGPT	182	0	0	0	21
singapore-sglt2	after	Claude	182	0	0	0	7
saudi-t1d	after	Perplexity	448	3	3	0	65
saudi-t1d	after	Gemini	448	1	1	0	411
saudi-t1d	after	ChatGPT	448	0	0	0	32
saudi-t1d	after	Claude	445	1	1	0	72

Text route: 0 qualifying answers in every series, engine and window (ledgerCounts.textQualifying 0).

Frozen and corrected either-route tallies (after window). In the registered markets, congress-derived sources were cited in 6 of 2,408 after-window answers; seven of 2,408 qualify under the either-route definition (one text-only answer found after the freeze: a UAE Gemini answer reporting the OCTANE result without a qualifying source, §71.5, §11.5.11(b)). The frozen tally and the corrected tally are shown separately. The frozen outputs are unchanged. OCTANE was added after registration (D3) and is not a registered readout.

Attribution under the corrected classification. The added answer is in the governed attribution population and named the medicine. Under the corrected classification, Gemini's UAE attribution table has two qualifying answers, both naming it, against 204 of 376 other scored answers (two-sided Fisher p = 0.503; minimum attainable p = 0.206): still structurally uninformative, and no verdict changes. The frozen calculation (frozen E1 classification: one qualifying answer, named, against 205 of 377; Fisher p = 1.000; minimum attainable p = 0.455) remains available as provenance.

Population	Answers	Frozen tally (either route)	Corrected tally (either route)
Registered arms (UAE and Singapore)	2,408	6 (0.249%)	7 (0.291%)
UAE, Gemini	420	1	2
Saudi Arabia (post-registration, D10)	1,789	5	5 (unchanged)
All three arms	4,197	11 (0.262%)	12 (0.286%)

10.3 E1 latency (verbatim)

Source: easd-2026-e1.md; latency[] in e1.json.

Series	Readout	Perplexity	Gemini	ChatGPT	Claude
uae-glp1	REIMAGINE 1 (CagriSema), LBA 11/12	not observed by day 12.1	not observed by day 12.1	not observed by day 12.1	not observed by day 12.1
uae-glp1	TRIUMPH-2 (retatrutide)	not observed by day 11.3	not observed by day 11.3	not observed by day 11.3	not observed by day 11.3
uae-glp1	CagriSema appetite, abstract 122	not observed by day 11.1	not observed by day 11.1	not observed by day 11.1	not observed by day 11.1
uae-glp1	ACHIEVE programme (orforglipron)	not observed by day 10.3	not observed by day 10.3	not observed by day 10.3	not observed by day 10.3
singapore-sglt2	SMARTEST (SGLT2 vs metformin first line)	not observed by day 10.3	not observed by day 10.3	not observed by day 10.3	not observed by day 10.3
uae-glp1	ACHIEVE-4 kidney outcomes, LBA 25	not observed by day 10	not observed by day 10	not observed by day 10	not observed by day 10
uae-glp1	Tirzepatide vs semaglutide proteomics, LBA 29	not observed by day 10	not observed by day 10	not observed by day 10	not observed by day 10
uae-glp1	OCTANE (switch to oral semaglutide, real-world), LBA 59 (post-registration, D3)	4.2 (UAE -C16)	6.2 (UAE -C18)	not observed by day 11.2	not observed by day 11.2
saudi-t1d	EASD 2026 type 1 diabetes programme (agent set + teplizumab abstracts) (post-registration, D10)	not observed by day 12.4	not observed by day 12.4	not observed by day 12.4	9.4 (SA -C17)
uae-glp1	Other EASD 2026 data (not in the readout set) (post-registration, E1 classes (c)/(d))	6.3 (UAE -C16)	8.3 (UAE -C18)	not observed by day 13.3	not observed by day 13.3
singapore-sglt2	Other EASD 2026 data (not in the readout set) (post-registration, E1 classes (c)/(d))	not observed by day 13.3	not observed by day 13.3	not observed by day 13.3	not observed by day 13.3
saudi-t1d	Other EASD 2026 data (not in the readout set) (post-registration, E1 classes (c)/(d))	7.4 (SA -C14)	9.4 (SA -C16)	not observed by day 13.4	10.4 (SA -C17)

Registered readouts: no qualifying citation or result text was verified by the implemented inspection procedure for any registered readout in its own series, battery and window. Results are reported as first observed detection or not detected within follow-up (§71.4).

10.4 E1 readout data reaching another arm (verbatim)

Source: crossArm[].

Readout	Registered series	Answer series	Engine	First cycle	Days from presentation	Content at citation
TRIUMPH-2 (retatrutide)	uae-glp1	saudi-t1d	Claude	SA-C17	8.4	unverified (page carries it today; not shown to have carried it when cited)

10.5 E1 qualifying sources (verbatim)

Source: easd-2026-e1.md "Ledger: qualifying sources"; sources[] in docs/evidence/easd-2026-e1-ledger.json.

Series	URL (resolved)	As cited	Engines	First observed in an answer	Class	Readouts	Page date	Fetchd	Qualifying sentence
saudi-t1d	https://diatribe.org/diabetes-research/top-diabetes-news-easd-2026	direct	anthropic	2026-10-08T16:31	d-media	EASD 2026 type 1 diabetes programme (agent set + teplizumab abstracts); TRIUMPH-2 (retatrutide); Other EASD 2026 data (not in the readout set)	2026-10-05	2026-10-11	New insights on diabetes treatment, management, and technology emerged from the European Association for the Study of Diabetes (EASD) 2026 conference in Milan, Italy.
saudi-t1d	[page cited only by Gemini: URL, publisher and title withheld, §13.3]	Gemini redirect	google	2026-10-07T16:32	c-sponsor	Other EASD 2026 data (not in the readout set)	2026-10-05	2026-10-10	[quotation withheld: page cited only by Gemini]
saudi-t1d	https://sadanews.ps/en/news/339566.html	direct	perplexity	2026-10-05T16:32	d-media	Other EASD 2026 data (not in the readout set)	not printed	2026-10-10	The study's results, led by the University Medical Center in the German city of Ulm in collaboration with Boston Children's Hospital and Harvard Medical School in the United States, were presented on Friday at the annual meeting of the European Association for the Study of Diabetes, currently taking place in Milan, Italy.
uae-glp1	https://biopharmadive.com/news/3-takeaways-from-european-obesity-scientific-meeting/831870	direct	perplexity	2026-10-04T14:02	d-media	OCTANE (switch to oral semaglutide, real-world), LBA 59; Other EASD 2026 data (not in the readout set)	2026-10-01	2026-10-10	At EASD, it presented data from a "real world" study of de-identified patient data it did in conjunction with direct-to-consumer partner Ro.
uae-glp1	https://biospace.com/press-releases/lillys-zepbound-tirzepatide-10-mg-and-15-mg-was-associated-with-more-weight-loss-than-wegovy-hd-semaglutide-injection-7-2-mg-in-a-new-indirect-treatment-comparison	direct	perplexity	2026-10-04T14:02	c-sponsor	Other EASD 2026 data (not in the readout set)	2026-09-29	2026-10-10	1,2,3 These results, presented at the 62nd Annual Meeting of the European Association for the Study of Diabetes (EASD) in Milan, Italy, add to the growing body of evidence comparing Zepbound and Wegovy, including SURMOUNT-5, a randomized controlled Phase 3b trial in which Zepbound 10 mg and 15 mg demonstrated superior weight loss compared with Wegovy 1.7 mg and 2.4 mg.
uae-glp1	https://biospace.com/press-releases/people-switching-from-weight-loss-injections-to-novos-wegovy-pill-lost-4-1-body-weight-on-average-after-three-months-according-to-real-world-data-presented-at-easd2026	direct	perplexity	2026-10-04T14:02	c-sponsor	OCTANE (switch to oral semaglutide, real-world), LBA 59	2026-09-30	2026-10-10	1 OCTANE was presented during a short oral session at the 62nd Annual Meeting of the European Association for the Study of Diabetes (EASD) in Milan, Italy, on September 30, 2026.
uae-glp1	https://cnbc.com/2026/09/30/eli-lilly-obesity-drug-amylin-tirzepatide-weight-loss-trial.html	direct	perplexity	2026-10-04T14:02	d-media	Other EASD 2026 data (not in the readout set)	2026-09-30	2026-10-10	A potential winning approach Results were presented at the European Association for the Study of Diabetes in Milan, Italy.
uae-glp1	[page cited only by Gemini: URL, publisher and title withheld, §13.3]	Gemini redirect	google	2026-10-06T14:04	d-media	Other EASD 2026 data (not in the readout set)	2026-10-05	2026-10-10	[quotation withheld: page cited only by Gemini]
uae-glp1	[page cited only by Gemini: URL, publisher and title withheld, §13.3]	Gemini redirect	google	2026-10-06T14:04	d-media	OCTANE (switch to oral semaglutide, real-world), LBA 59	2026-10-05	2026-10-10	[quotation withheld: page cited only by Gemini]
uae-glp1	[page cited only by Gemini: URL, publisher and title withheld, §13.3]	Gemini redirect	google	2026-10-06T14:04	d-media	OCTANE (switch to oral semaglutide, real-world), LBA 59	2026-09-30	2026-10-10	[quotation withheld: page cited only by Gemini]

10.6 E2: engines' naming of semaglutide (Wegovy) in the UAE and teplizumab (Tzield) in Saudi Arabia, against the bridged band (verbatim)

Source: docs/evidence/easd-2026-analysis.md; e2[] in easd-2026-analysis.json. Columns:

- **Before (ref):** governed v4 pooled K/N over the band-defining cycles.
- **Magnitude:** l2.
- **Beyond band:** pp beyond the nearer bridged edge, 0 when inside.
- **E1 answers (named) / Others (named):** the attribution split (§7.2.3).

Series	Engine	Before (ref)	During	After	After, question-clustered 95% (post-reg.)	Bridged band	Movement	Magnitude	Beyond band	E1 answers (named)	Others (named)	Fisher p	Verdict
uae-glp1	Perplexity	193/371 (52.0%)	22/54 (40.7%)	243/378 (64.3%)	56.6%–72.0%	42.0%–64.4%	no	+12.3pp DECISION-RELEVANT	0pp	1 (1)	377 (242)	1	congress sources without movement
uae-glp1	Gemini	416/701 (59.3%)	35/54 (64.8%)	206/378 (54.5%)	45.5%–63.8%	35.1%–78.4%	no	-0.5pp SMALL	0pp	1 (1)	377 (205)	1	congress sources without movement
uae-glp1	ChatGPT	212/702 (30.2%)	15/54 (27.8%)	110/378 (29.1%)	18.8%–39.1%	14.2%–46.9%	no	+0.8pp SMALL	0pp	0 (0)	378 (110)	—	neither
uae-glp1	Claude	336/702 (47.9%)	31/54 (57.4%)	256/378 (67.7%)	57.1%–78.0%	34.5%–67.3%	yes	+16.8pp MAJOR	0.4pp	0 (0)	378 (256)	—	movement without congress sources
uae-glp1	Four engines	1273/2794 (45.6%)	103/216 (47.7%)	815/1512 (53.9%)	47.2%–60.9%	36.9%–54.0%	no	+9.1pp POTENTIALLY RELEVANT	0pp	—	—	—	—
saudi-t1d (post-reg.)	Perplexity	58/589 (9.8%)	0/0 (n/a)	26/440 (5.9%)	2.9%–9.1%	3.8%–23.6%	no	-4.5pp SMALL	0pp	3 (1)	437 (25)	0.1674	congress sources without movement
saudi-t1d (post-reg.)	Gemini	54/603 (9.0%)	0/0 (n/a)	40/440 (9.1%)	4.3%–14.5%	3.0%–30.3%	no	-4.7pp SMALL	0pp	1 (0)	439 (40)	1	congress sources without movement
saudi-t1d (post-reg.)	ChatGPT	105/605 (17.4%)	0/0 (n/a)	81/440 (18.4%)	9.8%–27.7%	6.7%–35.1%	no	+1.1pp SMALL	0pp	0 (0)	440 (81)	—	neither
saudi-t1d (post-reg.)	Claude	30/605 (5.0%)	0/0 (n/a)	28/437 (6.4%)	3.2%–10.3%	-1.2%–13.3%	no	+0.8pp SMALL	0pp	0 (0)	437 (28)	—	neither
saudi-t1d (post-reg.)	Four engines	247/2402 (10.3%)	0/0 (n/a)	175/1757 (10.0%)	6.4%–13.9%	6.5%–23.3%	no	-1.8pp SMALL	0pp	—	—	—	—

Exact after-window rates and signed distances are in §6.4.

Registered verdicts, UAE.

- Claude: 256/378 = 67.725%, beyond the bridged upper edge by +0.395pp (+0.441pp on the exact-cycle band). Movement declared, MAJOR (+16.8pp). No E1-qualifying answer in Claude's after-window governed answers; verdict "movement without congress sources".
- Perplexity: 243/378 = 64.286%, inside by 0.074pp. No movement.
- Four engines: 815/1512 = 53.902%, inside by 0.078pp. No movement.
- Gemini and ChatGPT: well inside their bands.
- The question-clustered 95% interval contains the nearer edge for Claude, Perplexity and the four-engine rate. Those verdicts are borderline. They stand as registered and are not reversed.

Sensitivity of the Claude verdict. The pooled movement verdict is sensitive to the band-rule interpretation (\$6.3, \$11.5.2). A post hoc cycle-level analysis estimates a Claude increase of 16.8 percentage points (95% interval 7.2 to 26.4) under independent-cycle and bridge assumptions; it supports a descriptive increase within this design (\$11.5.3).

10.7 Singapore E2: positioning, descriptive (I4) (verbatim)

Source: singaporePositioning[]. Average rank to two decimals as printed; the JSON holds full precision.

Cycle	Date	Window	Average rank	Top competitor mention rates
SG-C01	2026-08-22	before-v4	1.72	Invokana 98.7%, Farxiga 98.7%, Forxiga 90.7%
SG-C02	2026-08-24	before-v4	1.74	Farxiga 100.0%, Invokana 98.7%, Forxiga 90.8%
SG-C03	2026-08-26	before-v4	1.7	Invokana 100.0%, Farxiga 100.0%, Forxiga 92.1%
SG-C04	2026-08-28	before-v4	1.79	Farxiga 100.0%, Invokana 97.4%, Forxiga 90.8%
SG-C05	2026-08-30	before-v4	1.72	Farxiga 98.7%, Invokana 97.4%, Forxiga 85.5%
SG-C06	2026-09-04	before-v4	1.72	Invokana 100.0%, Farxiga 100.0%, Forxiga 80.3%
SG-C07	2026-09-06	before-v4	1.64	Invokana 100.0%, Farxiga 100.0%, Forxiga 89.5%
SG-C08	2026-09-08	before-v4	1.63	Invokana 97.4%, Farxiga 97.4%, Forxiga 90.8%
SG-C09	2026-09-10	before-v4	1.64	Invokana 100.0%, Farxiga 100.0%, Forxiga 88.2%
SG-C10	2026-09-13	before-v4	1.69	Farxiga 98.7%, Invokana 97.4%, Forxiga 85.5%
SG-C11	2026-09-15	before-v4	1.8	Invokana 100.0%, Farxiga 100.0%, Forxiga 85.5%
SG-C12	2026-09-17	before-v4	1.61	Invokana 100.0%, Farxiga 100.0%, Forxiga 80.3%
SG-C13	2026-09-19	before-v4	1.79	Invokana 100.0%, Farxiga 100.0%, Forxiga 89.3%
SG-C14	2026-09-21	before-v4	1.59	Farxiga 98.7%, Invokana 97.3%, Forxiga 82.7%
SG-C15	2026-09-26	before-v4	1.64	Invokana 100.0%, Farxiga 100.0%, Forxiga 86.8%
SG-C17	2026-10-02	during	1.68	Invokana 100.0%, Farxiga 100.0%, Forxiga 88.2%
SG-C18	2026-10-05	after	1.54	Invokana 100.0%, Farxiga 100.0%, Forxiga 90.8%
SG-C19	2026-10-06	after	1.46	Invokana 100.0%, Farxiga 100.0%, Forxiga 81.6%
SG-C20	2026-10-07	after	1.67	Invokana 100.0%, Farxiga 100.0%, Forxiga 86.8%
SG-C21	2026-10-08	after	1.67	Invokana 100.0%, Farxiga 100.0%, Forxiga 84.2%
SG-C22	2026-10-09	after	1.55	Invokana 100.0%, Farxiga 100.0%, Forxiga 82.9%
SG-C23	2026-10-10	after	1.58	Invokana 100.0%, Farxiga 100.0%, Forxiga 88.2%
SG-C24	2026-10-11	after	1.57	Invokana 100.0%, Farxiga 100.0%, Forxiga 80.3%

10.8 E3 (UAE): investigational agents named, by window (verbatim)

Source: e3.namingsByWindow, e3.codesByWindow, e3.clinicallyMaterialByWindow, e3.strictSensitivity.

Window	Perplexity	Gemini	ChatGPT	Claude	All	Status correct / absent / wrong	Clinically material	Material, strict (post-reg.)
before-v4	1/765	3/779	4/780	2/780	10/3104	9 / 1 / 0	1	0
before-v5	0/58	0/60	1/60	0/60	1/238	1 / 0 / 0	0	0
during	0/60	0/60	0/60	0/60	0/240	0 / 0 / 0	0	0
after	5/420	1/420	0/420	4/420	10/1680	8 / 2 / 0	2	0

All 21 in-scope naming answers. Source: e3.rows in easd-2026-final-coding.json. Spans are answer text, verbatim, as stored (truncated by the coder at a fixed width). Prompt ids are omitted. Codes in the after window are provisional (status carried forward from 2 Oct).

Cycle	Date	Window	v	Engine	Which-medicine Q	Agent	Market status	Code	Qualifying span (answer text)	Note
UAE-C01	2026-08-22	before-v4	4	Gemini	no	retatrutide	not approved (record as of 2026-10-02)	status-correct	[Gemini text not reproduced, §13.3] (code kept)	—
UAE-C02	2026-08-24	before-v4	4	ChatGPT	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"An investigational triple agonist targeting GLP-1, GIP, and glucagon receptors. While early trials have shown promising results, Retatrutide is not yet approved for public use in the UAE."	Listed as an option, with status explicitly stated (ambiguity A3).
UAE-C02	2026-08-24	before-v4	4	ChatGPT	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"including Retatrutide, which remains under clinical investigation and is not authorized for use in the UAE."	—
UAE-C03	2026-08-26	before-v4	4	Gemini	no	retatrutide	not approved (record as of 2026-10-02)	status-correct	[Gemini text not reproduced, §13.3] (code kept)	—

Cycle	Date	Window	v	Engine	Which-medicine Q	Agent	Market status	Code	Qualifying span (answer text)	Note
UAE - C04	2026-08-28	before-v4	4	ChatGPT	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"An investigational triple agonist ... Retatrutide is currently under study and not yet approved for public use in the UAE. It is not legally available for prescription or purchase."	Listed as an option, with status explicitly stated (ambiguity A3).
UAE - C05	2026-08-30	before-v4	4	Claude	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"Retatrutide, an experimental injectable treatment for weight loss, ... remains under clinical investigation and has not been approved by the Emirates Drug Establishment for marketing or medical use within the UAE."	—
UAE - C08	2026-09-11	before-v4	4	Gemini	no	retatrutide	not approved (record as of 2026-10-02)	status-correct	[Gemini text not reproduced, §13.3] (code kept)	Correct for retatrutide; the same sentence lumps EDE-registered Ozempic and Mounjaro with it as unapproved (ambiguity A4: a status error about an APPROVED product is outside E3).
UAE - C08	2026-09-11	before-v4	4	Perplexity	no	retatrutide	not approved (record as of 2026-10-02)	status-correct	"Retatrutide has been reported as still in clinical-trial stage and not approved for use or distribution in the UAE."	—
UAE - C11	2026-09-21	before-v4	4	ChatGPT	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"peptide-based treatments like Retatrutide, which remain under clinical investigation and are not authorized for use in the UAE."	—
UAE - C13	2026-09-25	before-v4	4	Claude	yes	cagrisema	not approved (record as of 2026-10-02)	status-absent	"A trial showed Novo's next-generation weight-loss drug CagriSema failed to match Lilly's Zepbound's efficacy."	'Next-generation drug' and 'a trial showed' imply development but state no status (S3; ambiguity A2). Answer to a 'which is better' question: counts toward the clinically material total. Named as a comparator, not offered as an option.
UAE - C14	2026-09-27	before-v5	5	ChatGPT	yes	retatrutide	not approved (record as of 2026-10-02)	status-correct	"enforcement actions against unapproved weight-loss products, including peptide-based supplements like Retatrutide, which lack established safety profiles"	Status stated ('unapproved'); the answer calls the agent a 'supplement' (ambiguity A5: correct status, wrong product class).
UAE - C16	2026-10-04	after	5	Perplexity	no	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"proved by the EDE.[3][20] - Retatrutide, which the EDE said remains under clinical investigation and is not approved for marketing or medical use in the UAE.[3][20] If you mean **non-prescription "	—
UAE - C19	2026-10-07	after	5	Claude	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"online, including Retatrutide, a weight-loss injection that has not been approved for use or distribution in the UAE - Specialists have warned against purchasing weight-loss medicines through soci"	Named only to warn against buying it.
UAE - C20	2026-10-08	after	5	Perplexity	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-absent	"e, compounded GLP-1 products, or social-media injections.** Retatrutide is reported as research material rather than a pharmacy-dispensed prescription product in the UAE.[1][4] - **Do not switch between Rybelsus, Ozempic, Wegovy, and Mounja"	Ambiguity A10 (new in the after window): "research material", sold outside pharmacies, and a warning not to buy it; no statement that it is investigational, in trials or unapproved, so S3 codes it absent (as A2 did for "next-generation"). Answer to a which-medicine question, so it counts toward the clinically material total although it does not offer the agent.
UAE - C20	2026-10-08	after	5	Claude	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"Retatrutide remains in the clinical trial stage and has not been approved for use or distribution in the UAE . ## Recommendation: For UAE consumers seeking weight management with the best"	—
UAE - C22	2026-10-10	after	5	Perplexity	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"e, tirzepatide, retatrutide, or imported injections online; investigational products are not equivalent to pharmacy-dispensed medicines.[2][3] - Confirm the exact **dose , monthly quantity, cold-chain storage, and to"	Named only to warn against buying it; "investigational" stated.
UAE - C22	2026-10-10	after	5	Claude	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"The authority clarified that Retatrutide remains in the clinical trial stage and has not been approved for use or distribution in the UAE. --- ## Recommendation for UAE Consumers: ** "	—
UAE - C22	2026-10-10	after	5	Gemini	no	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	[Gemini text not reproduced, §13.3] (code kept)	—
UAE - C23	2026-10-11	after	5	Perplexity	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"A valid prescription from a UAE-licensed clinician is required for these treatments, and products such as retatrutide remain investigational rather than ordinary pharmacy options.[3][4] Rybelsus, Ozempic and Wegovy all involve semaglutide b"	Named only to warn against buying it; "investigational" stated.
UAE - C23	2026-10-11	after	5	Perplexity	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-absent	" Do not treat online "peptide" products as alternatives: Retatrutide and similar research compounds are not established UAE pharmacy treatments; use licensed pharmacies and prescribed products only.[2][3] ## UAE cost and ac "	Ambiguity A10 (as on 8 Oct, answer 95eb4702): "research compounds" that are "not established UAE pharmacy treatments", with a warning not to treat them as alternatives; no statement that it is investigational, in trials or unapproved, so S3 codes it absent. Answer to a which-medicine question, so it counts toward the clinically material total although it does not offer the agent.
UAE - C23	2026-10-11	after	5	Claude	yes	retatrutide	not approved as of 2026-10-02; later status not established	status-correct	"The authority clarified that Retatrutide remains in the clinical trial stage and has not been approved for use or distribution in the UAE . --- ## Consumer Recommendation: **For the best safety profile as a UAE consum "	Named only to report the EDE warning.

The two after-window "clinically material" answers are both Perplexity answers to which-medicine questions. Both name retatrutide only to warn readers away from it. Neither states that it is investigational, in trials or unapproved, which is why S3 codes them *status absent*:

- 8 Oct: "Retatrutide is reported as research material rather than a pharmacy-dispensed prescription product in the UAE".
- 11 Oct: "Retatrutide and similar research compounds are not established UAE pharmacy treatments; use licensed pharmacies and prescribed products only".

They are not presented as options and are not clinical errors in their wording. The strict count, which removes ambiguity-labelled codes, is 0.

10.9 E5 (Singapore, empagliflozin (Jardiance) series): first-line framing, by window (verbatim)

Source: e5[window].

Window	Answers	Candidates	General	High-risk	With metformin (reported, not counted)	Not first-line	Protocol first-line	Strict (post-reg.)
before-v4	1554	202	0	1	48	153	1	0
before-v5	0	0	0	0	0	0	0	0
during	104	9	0	0	4	5	0	0
after	728	53	0	1	11	41	1	0

10.10 E4 (UAE, semaglutide (Wegovy) series): hand-verified counts only (quarantined, D17)

Source: e4[window] in easd-2026-analysis.json. The label and published columns are the script's attribution, which failed audit, and are not reproduced here (they remain in the file, marked "Not for publication as coded"). The two columns shown are hand-checked under D4. Answers carrying a figure in either category, by window:

Window	Answers	Answers with a hand-checked figure of no identifiable source	Answers with a hand-checked congress-window figure
before-v4	3104	62	0
before-v5	238	2	0
during	240	3	0
after	1680	20	2

10.11 Saudi Arabia (post-registration, D10) (verbatim)

Source: saudi.e3NamingsByWindow, saudi.opportunity, saudi.e3iExploratory. Since D21 item 1, the Saudi E3 table states its namings, its denominator ("Answers searched") and its opportunity count over the same set: every captured answer of the admitted cycles, focused questions included. The governed E2 population is shown for reference only.

Window	Namings	Answers searched	Opportunity answers (within those searched)	Governed E2 population (reference)
before-v4	0	2446	44	2402
before-v5	0	0	0	0
during	0	0	0	0
after	0	1789	32	1757

E3-i (exploratory). Denominator: sampled namings coded as a type 1 diabetes use (stated, absent or wrong); namings that only rule the molecule out are excluded.

Window	Namings (sampled)	T1D-use namings (distinct answers)	No status statement	Share	Question-clustered 95%	Strict share	In which-medicine answers
before-v4	112 (91)	56 (45)	38	67.9%	44.4%–87.3%	67.9%	8
before-v5	0 (0)	0 (0)	0	n/a	n/a	n/a	0
during	0 (0)	0 (0)	0	n/a	n/a	n/a	0
after	206 (120)	39 (31)	25	64.1%	40.6%–82.0%	64.1%	8

Glycaemic-use recode (post hoc, D24, \$11.5.9). In the exploratory Saudi-arm sensitivity recode, 30/44 before-window and 16/18 after-window sampled glycaemic-use namings lacked qualifying indication-status wording under the study's rules. Saudi product-specific labeling was not independently established; the finding identifies a communication issue for regulatory review. The recode was made by the same model family as the other coders.

E3 by agent, all windows: 0 namings for each of the ten agents (headline.e3.byAgent).

11. Sensitivity and robustness

11.1 Strict E3 and E5 (D16)

Endpoint	Window	Registered count	Strict count
E3 UAE clinically material	before-v4	1	0
E3 UAE clinically material	after	2	0
E5 protocol first-line	before-v4	1	0
E5 protocol first-line	after	1	0
Saudi E3-i share without a status statement	before-v4	67.9%	67.9%
Saudi E3-i share without a status statement	after	64.1%	64.1%

Sources: e3.strictSensitivity, e5[w].protocolFirstLineStrict, saudi.e3iExploratory[w].strictShare. Every code removed by the strict scenario carries an ambiguity note (A2, A6 or A10).

11.2 Full-precision E2 (D19)

See §6.4. Under the exact-cycle band every verdict is the same as the registered verdict (movementUnderExactCycleBand equals movementDeclared in all 10 rows). Rows whose clustered 95% interval contains the nearer edge:

- UAE: Perplexity, Claude, four engines.
- Saudi: Perplexity, four engines.

11.3 Clustered intervals

The question-clustered intervals are in the E2 table (§10.6) and the Saudi E3-i table (§10.11). They widen the after-window uncertainty because answers to the same question are correlated across cycles and engines. They are robustness only. The registered rule compares a pooled rate with a band and does not use them.

11.4 E1 with and without post-registration readouts

The *qualification* of an answer (citing a class a–d source, or text reporting a readout) is registered. The readout labels beyond the registered seven are post-registration.

Series	Engine	Qualifying after-window answers, all labels	Carrying a registered readout (own series)	OCTANE (D3)	Saudi programme (D10)	Other EASD 2026 data only
saudi-t1d	Claude	1	0	0	1	0
saudi-t1d	Gemini	1	0	0	0	1
saudi-t1d	Perplexity	3	0	0	0	3
uae-glp1	Gemini	1	0	1	0	0
uae-glp1	Perplexity	5	0	3	0	2

Removing every post-registration readout leaves the registered latency table unchanged: all 28 registered engine × readout cells read "not observed by day N".

11.5 Post-registration robustness analyses (D24)

Status.

- These analyses were computed after the results were seen, by a separate script (`scripts/easd-2026-study/robustness.ts`), into separate outputs:
 - `docs/evidence/easd-2026-robustness.md` and `docs/evidence/computed/easd-2026-robustness.json`;
 - a per-answer outcome table, `docs/evidence/computed/easd-2026-per-answer-outcomes.csv`;
 - Saudi use-context codes, `docs/evidence/easd-2026-saudi-e3i-use-context.json`.
- The script first reproduces every registered count and verdict and stops on any mismatch.
- **No registered verdict changes. Nothing here replaces a registered figure.**
- **§11.5.1 is the exception:** it reports the three descriptive source analyses that protocol §5 registered and the frozen analysis never ran. It is **registered, reported late**.
- Everything from §11.5.2 on is **post-registration robustness**.

The local page that §11.5.1(b) and §11.5.12 discuss is identified only by its neutral type: a local commercial blog post (UAE), August 2026, before the congress. Its name, domain, URL and content are withheld (§10, D22; a finding of the second-round internal claim-to-evidence check, §9.7). Its two sentences are referred to only as S1, a class statement, and S2, an efficacy-figure sentence, and are matched by hashes.

A page displaying an August 2026 date was fetched on 11 October and first appeared in Claude's resolved study citations on 29 September. Historical first publication and the complete page content at each citation date were not independently archived. Shared wording supports an association between the page and the answers; it does not establish exclusive source provenance or causation.

Everything below, from §11.5.1 onwards, is copied from `docs/evidence/easd-2026-robustness.md` by the generator. The heading levels and numbers are changed, and 26 passages are reworded for publication: internal review references are replaced by the issue and its protocol section or deviation, and citation, interval, latency, attribution-classification, indication-wording and page-provenance statements are qualified. No count is changed; the shares in the §11.5.1(c) source-mix tables are recomputed from their integer counts (see the note there). The Gemini answers' as-written figure in §11.5.11(b) is withheld.

11.5.1 Registered §5 descriptive source analyses — registered §5, reported late under D24

Protocol §5 registered three descriptive source analyses. The frozen analysis did not report them and no deviation recorded the omission (the omission concerned protocol §5 and is recorded under D24). They are reported here, late, as registered descriptives.

11.5.1(a) The Note 04 clinic page: Perplexity citations per cycle

Answers citing the page (exact normalised URL) over the engine's captured answers in each admitted UAE cycle. The page and its publisher are not named. Gemini's page-level match needs resolved redirect links, which exist for the congress and after windows only.

Cycle date	Window	Perplexity citing / answers	Perplexity citing and naming	Claude	ChatGPT	Gemini
2026-08-22	before-v4	0/54 (0.0%)	0	0/60	0/60	n/a
2026-08-24	before-v4	0/60 (0.0%)	0	0/60	0/60	n/a
2026-08-26	before-v4	0/60 (0.0%)	0	0/60	0/60	n/a
2026-08-28	before-v4	0/60 (0.0%)	0	0/60	0/60	n/a
2026-08-30	before-v4	0/59 (0.0%)	0	0/60	0/60	n/a
2026-09-04	before-v4	8/60 (13.3%)	8	0/60	0/60	n/a
2026-09-09	before-v4	9/60 (15.0%)	9	0/60	0/60	n/a
2026-09-11	before-v4	8/60 (13.3%)	8	0/60	0/60	n/a
2026-09-14	before-v4	7/57 (12.3%)	7	0/60	0/60	n/a
2026-09-16	before-v4	12/56 (21.4%)	11	0/60	0/60	n/a
2026-09-21	before-v4	10/59 (17.0%)	10	0/60	0/60	n/a
2026-09-23	before-v4	7/60 (11.7%)	7	0/60	0/60	n/a
2026-09-25	before-v4	4/60 (6.7%)	4	0/60	0/60	n/a
2026-09-27	before-v5	3/58 (5.2%)	3	0/60	0/60	n/a
2026-09-29	during	3/60 (5.0%)	3	0/60	0/60	0/60
2026-10-04	after	4/60 (6.7%)	4	0/60	0/60	0/60

Cycle date	Window	Perplexity citing / answers	Perplexity citing and naming	Claude	ChatGPT	Gemini
2026-10-06	after	7/60 (11.7%)	4	0/60	0/60	0/60
2026-10-07	after	3/60 (5.0%)	3	0/60	0/60	0/60
2026-10-08	after	1/60 (1.7%)	1	0/60	0/60	0/60
2026-10-09	after	2/60 (3.3%)	2	0/60	0/60	0/60
2026-10-10	after	3/60 (5.0%)	3	0/60	0/60	0/60
2026-10-11	after	4/60 (6.7%)	4	0/60	0/60	0/60

What this shows. The page first appears in Perplexity's citations in the 2026-09-04 cycle, peaked at 12 answers in one cycle, and appears in 7 of 7 after-window cycles (4, 7, 3, 1, 2, 3, 4 answers). **What it does not show.** Whether the page caused any naming change: citation is a link, not an effect, and the counts are not compared with a band.

11.5.1(b) New local or regional pages drawing ≥ 10% of one engine's answers in a cycle

Congress- and after-window cycles. A page is NEW when no before-window cycle of the same series cited it (E1's manifest; Gemini's pre-congress redirects were never resolved, so Gemini over-cludes, D16). LOCAL OR REGIONAL: the host is under a UAE, Saudi, Singapore or other GCC suffix, or is classified clinic or consumer/retail. A page is listed when the answers citing it are ≥ 10% of that engine's captured answers in the cycle. Pages are typed and given an opaque id, never named.

ChatGPT, Claude and Perplexity (every listed page-cycle):

Series	Cycle date	Engine	Page (opaque id)	Type	Answers citing / answers	Share	Naming among citing
saudi-t1d	2026-10-05	ChatGPT	P17	consumer / retail (incl. pharmacy) page, non-regional domain	7/56	12.5%	1/7
singapore-sglt2	2026-10-02	Perplexity	P24	consumer / retail (incl. pharmacy) page, non-regional domain	7/26	26.9%	7/7
singapore-sglt2	2026-10-02	Perplexity	P25	consumer / retail (incl. pharmacy) page, non-regional domain	5/26	19.2%	5/5
singapore-sglt2	2026-10-02	Perplexity	P26	other (classified) page, Singapore domain (public: data.gov.sg)	4/26	15.4%	4/4
singapore-sglt2	2026-10-02	Perplexity	P27	clinic / hospital / care provider page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-05	Perplexity	P24	consumer / retail (incl. pharmacy) page, non-regional domain	7/26	26.9%	7/7
singapore-sglt2	2026-10-05	Perplexity	P35	clinic / hospital / care provider page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-05	Perplexity	P36	regulator / government health body page, Singapore domain (public: hsa.gov.sg)	3/26	11.5%	3/3
singapore-sglt2	2026-10-05	Perplexity	P37	unclassified page, Singapore domain	3/26	11.5%	3/3
singapore-sglt2	2026-10-06	Perplexity	P24	consumer / retail (incl. pharmacy) page, non-regional domain	7/26	26.9%	7/7
singapore-sglt2	2026-10-06	Perplexity	P35	clinic / hospital / care provider page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-06	ChatGPT	P42	consumer / retail (incl. pharmacy) page, non-regional domain	3/26	11.5%	3/3
singapore-sglt2	2026-10-06	ChatGPT	P43	consumer / retail (incl. pharmacy) page, non-regional domain	3/26	11.5%	3/3
singapore-sglt2	2026-10-07	ChatGPT	P45	consumer / retail (incl. pharmacy) page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-09	ChatGPT	P42	consumer / retail (incl. pharmacy) page, non-regional domain	3/26	11.5%	3/3
singapore-sglt2	2026-10-10	ChatGPT	P45	consumer / retail (incl. pharmacy) page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-10	ChatGPT	P43	consumer / retail (incl. pharmacy) page, non-regional domain	4/26	15.4%	4/4
singapore-sglt2	2026-10-11	ChatGPT	P45	consumer / retail (incl. pharmacy) page, non-regional domain	3/26	11.5%	3/3
uae-ghp1	2026-09-29	Perplexity	P50	regulator / government health body page, UAE domain (public: doh.gov.ae)	7/60	11.7%	1/7
uae-ghp1	2026-09-29	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	12/60	20.0%	10/12
uae-ghp1	2026-10-04	Perplexity	P62	consumer / retail (incl. pharmacy) page, non-regional domain	15/60	25.0%	8/15
uae-ghp1	2026-10-04	Perplexity	P50	regulator / government health body page, UAE domain (public: doh.gov.ae)	9/60	15.0%	2/9
uae-ghp1	2026-10-04	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	22/60	36.7%	20/22
uae-ghp1	2026-10-06	Perplexity	P62	consumer / retail (incl. pharmacy) page, non-regional domain	13/60	21.7%	2/13
uae-ghp1	2026-10-06	Perplexity	P50	regulator / government health body page, UAE domain (public: doh.gov.ae)	9/60	15.0%	3/9
uae-ghp1	2026-10-06	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	23/60	38.3%	23/23
uae-ghp1	2026-10-07	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	24/60	40.0%	24/24
uae-ghp1	2026-10-08	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	21/60	35.0%	21/21
uae-ghp1	2026-10-09	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	20/60	33.3%	18/20
uae-ghp1	2026-10-10	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	23/60	38.3%	20/23
uae-ghp1	2026-10-11	ChatGPT	P85	consumer / retail (incl. pharmacy) page, non-regional domain	6/60	10.0%	0/6
uae-ghp1	2026-10-11	ChatGPT	P86	consumer / retail (incl. pharmacy) page, non-regional domain	6/60	10.0%	0/6
uae-ghp1	2026-10-11	Claude	P61	a local commercial blog post (UAE), August 2026, before the congress	19/60	31.7%	17/19

Gemini (summary; every page-cycle is in the JSON). Gemini's pre-congress redirect links were never resolved, so a page Gemini cited before the congress also counts as new here (D16): Gemini's rows over-include and are not comparable with the other engines'.

Series	Cycle date	Pages listed	Largest share	Types of the listed pages
saudi-t1d	2026-10-04	15	50.0%	consumer / retail (incl. pharmacy) page, non-regional domain ×11; consumer / retail (incl. pharmacy) page, Saudi domain ×4
saudi-t1d	2026-10-05	10	44.6%	consumer / retail (incl. pharmacy) page, non-regional domain ×7; consumer / retail (incl. pharmacy) page, Saudi domain ×3
saudi-t1d	2026-10-06	12	46.4%	consumer / retail (incl. pharmacy) page, non-regional domain ×9; consumer / retail (incl. pharmacy) page, Saudi domain ×3
saudi-t1d	2026-10-07	10	48.2%	consumer / retail (incl. pharmacy) page, non-regional domain ×8; consumer / retail (incl. pharmacy) page, Saudi domain ×2
saudi-t1d	2026-10-08	12	50.0%	consumer / retail (incl. pharmacy) page, non-regional domain ×9; consumer / retail (incl. pharmacy) page, Saudi domain ×3
saudi-t1d	2026-10-09	11	48.2%	consumer / retail (incl. pharmacy) page, non-regional domain ×8; consumer / retail (incl. pharmacy) page, Saudi domain ×2; clinic / hospital / care provider page, non-regional domain ×1
saudi-t1d	2026-10-10	11	60.7%	consumer / retail (incl. pharmacy) page, non-regional domain ×8; consumer / retail (incl. pharmacy) page, Saudi domain ×3

Series	Cycle date	Pages listed	Largest share	Types of the listed pages
saudi-t1d	2026-10-11	11	53.6%	consumer / retail (incl. pharmacy) page, non-regional domain ×8; consumer / retail (incl. pharmacy) page, Saudi domain ×3
singapore-sglit2	2026-10-02	7	46.2%	consumer / retail (incl. pharmacy) page, non-regional domain ×4; clinic / hospital / care provider page, non-regional domain ×2; clinic / hospital / care provider page, Singapore domain ×1
singapore-sglit2	2026-10-05	8	46.2%	clinic / hospital / care provider page, non-regional domain ×4; consumer / retail (incl. pharmacy) page, non-regional domain ×3; regulator / government health body page, Singapore domain ×1
singapore-sglit2	2026-10-06	5	23.1%	clinic / hospital / care provider page, non-regional domain ×3; consumer / retail (incl. pharmacy) page, non-regional domain ×2
singapore-sglit2	2026-10-07	8	50.0%	consumer / retail (incl. pharmacy) page, non-regional domain ×5; clinic / hospital / care provider page, non-regional domain ×3
singapore-sglit2	2026-10-08	6	38.5%	clinic / hospital / care provider page, non-regional domain ×3; consumer / retail (incl. pharmacy) page, non-regional domain ×2; clinic / hospital / care provider page, Singapore domain ×1
singapore-sglit2	2026-10-09	6	30.8%	clinic / hospital / care provider page, non-regional domain ×3; consumer / retail (incl. pharmacy) page, non-regional domain ×2; clinic / hospital / care provider page, Singapore domain ×1
singapore-sglit2	2026-10-10	9	53.8%	clinic / hospital / care provider page, non-regional domain ×6; consumer / retail (incl. pharmacy) page, non-regional domain ×2; clinic / hospital / care provider page, Singapore domain ×1
singapore-sglit2	2026-10-11	6	50.0%	clinic / hospital / care provider page, non-regional domain ×3; consumer / retail (incl. pharmacy) page, non-regional domain ×2; clinic / hospital / care provider page, Singapore domain ×1
uae-ghp1	2026-09-29	10	20.0%	clinic / hospital / care provider page, non-regional domain ×3; clinic / hospital / care provider page, UAE domain ×2; consumer / retail (incl. pharmacy) page, UAE domain ×2; consumer / retail (incl. pharmacy) page, non-regional domain ×2; unclassified page, UAE domain ×1
uae-ghp1	2026-10-04	13	21.7%	clinic / hospital / care provider page, UAE domain ×5; clinic / hospital / care provider page, non-regional domain ×4; consumer / retail (incl. pharmacy) page, non-regional domain ×2; consumer / retail (incl. pharmacy) page, UAE domain ×1; unclassified page, UAE domain ×1
uae-ghp1	2026-10-06	10	25.0%	consumer / retail (incl. pharmacy) page, non-regional domain ×4; clinic / hospital / care provider page, UAE domain ×2; consumer / retail (incl. pharmacy) page, UAE domain ×2; clinic / hospital / care provider page, non-regional domain ×1; insurer page, UAE domain ×1
uae-ghp1	2026-10-07	7	20.0%	clinic / hospital / care provider page, UAE domain ×3; consumer / retail (incl. pharmacy) page, UAE domain ×2; consumer / retail (incl. pharmacy) page, non-regional domain ×1; clinic / hospital / care provider page, non-regional domain ×1
uae-ghp1	2026-10-08	9	18.3%	clinic / hospital / care provider page, UAE domain ×4; clinic / hospital / care provider page, non-regional domain ×2; consumer / retail (incl. pharmacy) page, non-regional domain ×2; consumer / retail (incl. pharmacy) page, UAE domain ×1
uae-ghp1	2026-10-09	8	23.3%	consumer / retail (incl. pharmacy) page, non-regional domain ×3; clinic / hospital / care provider page, UAE domain ×3; clinic / hospital / care provider page, non-regional domain ×1; insurer page, UAE domain ×1
uae-ghp1	2026-10-10	10	13.3%	clinic / hospital / care provider page, non-regional domain ×5; consumer / retail (incl. pharmacy) page, non-regional domain ×2; clinic / hospital / care provider page, UAE domain ×2; consumer / retail (incl. pharmacy) page, UAE domain ×1
uae-ghp1	2026-10-11	7	18.3%	clinic / hospital / care provider page, non-regional domain ×2; clinic / hospital / care provider page, UAE domain ×2; consumer / retail (incl. pharmacy) page, non-regional domain ×1; unclassified page, UAE domain ×1; consumer / retail (incl. pharmacy) page, UAE domain ×1

Claude, UAE: page P61. A page new to the series (never cited in any before-window cycle, including the 27 Sep v5 cycle, among the resolved citations examined; pre-congress Gemini destinations were unresolved) first drew 12/60 of Claude's answers on 2026-09-29 and 22/60, 23/60, 24/60, 21/60, 20/60, 23/60, 19/60 in the after-window cycles; 153 of the 164 citing answers across these cycles name Wegovy. Its type, from the page's own content: a local commercial blog post (UAE), August 2026, before the congress. The type and date come from a read-only fetch on 11 Oct; item 12 tests textual overlap with the answers by echo. It is never named.

What this shows. Which newly cited local or regional pages carried a sizeable share of one engine's answers after the congress (UAE distinct pages listed: Perplexity 2, Claude 1, ChatGPT 2, Gemini 32 with over-inclusion). In the UAE, Claude's after-window citations concentrate on one new local page that entered on 29 Sep, the cycle in which its naming first rose. **What it does not show.** That any page moved naming (naming among citing answers is high partly because the page discusses the medicine); that a page said then what it says now; or anything about the publishers, who are never named.

11.5.1(c) Source-class mix per engine per window

Unit: distinct cited links per answer. Classifier: rule-based, by host, written in the script (named public entities; hand-assigned local businesses held as hashes; government and academic suffixes; host tokens). 4408 distinct hosts; 151,614 links processed by the classifier, including 5.4% unclassified. "Other" = reference publishers, professional or patient organisations, academic hosts, non-health government, market research, insurers, directories and social platforms (link counts by subclass in the JSON). Gemini is classified by the domain its API returns as each result's title.

Percentages are independently rounded to one decimal place and may not sum to 100%. The supplement's display is recomputed from the same integer counts (half-up rounding); the frozen robustness output, and its copy in the evidence package, keeps its own rounding as provenance (19 of 360 displayed shares differ from the frozen display, each by 0.1 point).

UAE, Wegovy (registered) — share of links by class (links)

Engine	Window	Answers	Regulator / government health body	Journal / literature / trial registry	Congress	Sponsor / manufacturer	News / medical media	Clinic / hospital / care provider	Consumer / retail (incl. pharmacy)	Other (classified)	Unclassified
Perplexity	before-v4	765	8.9% (1303)	5.7% (828)	0.0% (0)	0.0% (2)	11.3% (1642)	32.7% (4762)	28.3% (4121)	9.9% (1439)	3.3% (477)
Perplexity	before-v5	58	8.4% (94)	4.3% (48)	0.0% (0)	0.0% (0)	14.9% (168)	30.5% (343)	29.3% (329)	8.8% (99)	3.8% (43)
Perplexity	during	60	9.5% (112)	4.6% (54)	0.0% (0)	0.0% (0)	17.9% (212)	26.8% (318)	28.6% (339)	9.7% (115)	3.0% (35)
Perplexity	after	420	6.7% (361)	3.9% (211)	0.0% (0)	0.1% (3)	12.7% (683)	21.6% (1158)	19.0% (1018)	33.0% (1770)	3.1% (165)
Gemini	before-v4	779	4.9% (1030)	0.6% (125)	0.0% (0)	0.5% (105)	10.2% (2143)	40.7% (8510)	25.2% (5268)	9.6% (2017)	8.2% (1726)
Gemini	before-v5	60	3.8% (63)	0.6% (10)	0.0% (0)	0.5% (8)	9.6% (159)	41.7% (689)	24.5% (406)	10.6% (176)	8.6% (143)
Gemini	during	60	3.5% (54)	0.4% (6)	0.0% (0)	0.5% (8)	9.6% (149)	43.9% (683)	24.3% (378)	9.4% (146)	8.5% (132)
Gemini	after	420	4.5% (507)	0.5% (53)	0.0% (0)	0.7% (81)	10.1% (1132)	40.4% (4532)	24.2% (2719)	10.8% (1210)	8.9% (996)
ChatGPT	before-v4	780	2.5% (96)	0.2% (9)	0.0% (0)	0.2% (9)	8.8% (338)	24.3% (937)	42.9% (1653)	11.1% (427)	9.9% (383)
ChatGPT	before-v5	60	0.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)	8.1% (24)	30.8% (91)	44.4% (131)	10.2% (30)	6.1% (18)
ChatGPT	during	60	2.0% (6)	0.7% (2)	0.0% (0)	0.3% (1)	7.4% (22)	25.5% (76)	44.3% (132)	10.7% (32)	9.1% (27)

Engine	Window	Answers	Regulator / government health body	Journal / literature / trial registry	Congress	Sponsor / manufacturer	News / medical media	Clinic / hospital / care provider	Consumer / retail (incl. pharmacy)	Other (classified)	Unclassified
ChatGPT	after	420	1.8% (38)	0.2% (5)	0.0% (0)	0.3% (7)	9.6% (199)	22.4% (465)	45.8% (949)	11.5% (239)	8.2% (171)
Claude	before-v4	780	2.3% (140)	5.9% (358)	0.0% (0)	0.1% (4)	16.1% (978)	31.5% (1916)	22.2% (1350)	15.2% (922)	6.7% (406)
Claude	before-v5	60	2.1% (10)	5.1% (24)	0.0% (0)	0.0% (0)	16.6% (78)	24.5% (115)	26.7% (125)	18.3% (86)	6.6% (31)
Claude	during	60	2.5% (12)	5.9% (28)	0.0% (0)	0.0% (0)	14.7% (70)	26.0% (124)	24.9% (119)	18.4% (88)	7.5% (36)
Claude	after	420	2.2% (72)	6.8% (220)	0.0% (0)	0.0% (0)	14.5% (472)	26.2% (851)	29.4% (957)	16.3% (529)	4.7% (153)

Singapore, Jardiance (registered) — share of links by class (links)

Engine	Window	Answers	Regulator / government health body	Journal / literature / trial registry	Congress	Sponsor / manufacturer	News / medical media	Clinic / hospital / care provider	Consumer / retail (incl. pharmacy)	Other (classified)	Unclassified
Perplexity	before-v4	385	25.3% (1771)	10.3% (723)	0.0% (0)	0.8% (54)	2.2% (153)	34.2% (2391)	10.8% (753)	15.2% (1065)	1.3% (89)
Perplexity	during	26	19.8% (102)	8.3% (43)	0.0% (0)	1.2% (6)	16.1% (83)	24.0% (124)	8.7% (45)	14.7% (76)	7.2% (37)
Perplexity	after	182	30.9% (689)	6.6% (148)	0.0% (0)	0.8% (17)	7.6% (170)	29.0% (647)	7.6% (170)	13.0% (289)	4.4% (98)
Gemini	before-v4	389	15.4% (956)	3.4% (214)	0.0% (0)	2.2% (140)	3.5% (216)	37.7% (2344)	8.6% (536)	25.9% (1614)	3.3% (205)
Gemini	during	26	11.2% (54)	1.0% (5)	0.0% (0)	1.7% (8)	5.8% (28)	32.0% (154)	11.9% (57)	34.1% (164)	2.3% (11)
Gemini	after	182	13.6% (434)	2.4% (77)	0.0% (0)	1.0% (32)	4.5% (143)	36.5% (1164)	9.4% (300)	29.7% (946)	2.8% (90)
ChatGPT	before-v4	390	62.8% (997)	0.3% (4)	0.0% (0)	0.0% (0)	0.7% (11)	7.6% (120)	22.2% (353)	3.5% (55)	3.0% (47)
ChatGPT	during	26	61.5% (75)	0.0% (0)	0.0% (0)	0.0% (0)	3.3% (4)	7.4% (9)	23.8% (29)	3.3% (4)	0.8% (1)
ChatGPT	after	182	62.9% (461)	0.3% (2)	0.0% (0)	0.0% (0)	0.3% (2)	7.1% (52)	21.0% (154)	2.9% (21)	5.6% (41)
Claude	before-v4	390	23.1% (465)	8.8% (178)	0.0% (0)	1.1% (23)	0.5% (11)	13.1% (263)	25.2% (508)	24.7% (497)	3.5% (70)
Claude	during	26	16.9% (20)	6.8% (8)	0.0% (0)	0.8% (1)	0.0% (0)	21.2% (25)	28.0% (33)	20.3% (24)	5.9% (7)
Claude	after	182	27.7% (222)	7.9% (63)	0.0% (0)	0.9% (7)	0.1% (1)	20.1% (161)	16.9% (135)	21.5% (172)	5.0% (40)

Saudi Arabia, Tzield (post-registration, D10) — share of links by class (links)

Engine	Window	Answers	Regulator / government health body	Journal / literature / trial registry	Congress	Sponsor / manufacturer	News / medical media	Clinic / hospital / care provider	Consumer / retail (incl. pharmacy)	Other (classified)	Unclassified
Perplexity	before-v4	600	18.1% (2077)	43.0% (4938)	0.1% (9)	7.3% (836)	6.4% (740)	5.9% (681)	1.6% (184)	14.1% (1624)	3.4% (389)
Perplexity	after	448	19.9% (1194)	32.7% (1958)	0.0% (0)	5.9% (356)	8.6% (515)	2.6% (155)	0.5% (27)	26.9% (1611)	3.0% (177)
Gemini	before-v4	614	13.2% (1844)	6.5% (903)	0.0% (6)	8.6% (1209)	4.9% (692)	8.8% (1227)	29.3% (4094)	22.6% (3165)	6.1% (847)
Gemini	after	448	12.3% (1240)	4.8% (486)	0.0% (4)	7.7% (779)	4.8% (481)	9.6% (971)	34.5% (3492)	20.3% (2053)	6.0% (609)
ChatGPT	before-v4	616	10.2% (183)	16.9% (303)	0.0% (0)	2.3% (42)	4.5% (80)	8.4% (151)	37.3% (668)	10.5% (187)	9.8% (175)
ChatGPT	after	448	13.2% (171)	17.6% (228)	0.0% (0)	1.5% (19)	2.7% (35)	6.8% (88)	40.1% (518)	10.0% (129)	8.0% (104)
Claude	before-v4	616	1.4% (61)	29.8% (1321)	0.0% (1)	1.6% (71)	9.6% (425)	2.8% (123)	3.1% (136)	48.1% (2131)	3.7% (164)
Claude	after	445	1.6% (50)	32.9% (1024)	0.0% (0)	2.2% (69)	9.1% (284)	3.7% (114)	4.0% (125)	43.5% (1353)	2.9% (89)

Claude, UAE: 27 Sep (v5, before) against 29 Sep (congress window)

Governed naming: 22/54 (40.7%) on 2026-09-27, 31/54 (57.4%) on 2026-09-29. Answers with at least one citation: 57/60 and 57/60. Distinct domains: 173 and 175 (64 new in the later cycle, 62 absent from it).

Class	Earlier: links (share)	Later: links (share)	Change (pp)	Earlier: answers citing (naming)	Later: answers citing (naming)
Regulator / government health body	10 (2.1%)	12 (2.5%)	+0.4	9 (5)	11 (8)
Journal / literature / trial registry	24 (5.1%)	28 (5.9%)	+0.8	12 (4)	14 (6)
Congress	0 (0.0%)	0 (0.0%)	0.0	0 (0)	0 (0)
Sponsor / manufacturer	0 (0.0%)	0 (0.0%)	0.0	0 (0)	0 (0)
News / medical media	78 (16.6%)	70 (14.7%)	-1.9	35 (26)	31 (24)
Clinic / hospital / care provider	115 (24.5%)	124 (26.0%)	+1.5	39 (25)	40 (30)
Consumer / retail (incl. pharmacy)	125 (26.7%)	119 (24.9%)	-1.7	44 (18)	42 (23)
Other (classified)	86 (18.3%)	88 (18.4%)	+0.1	43 (17)	44 (27)
Unclassified	31 (6.6%)	36 (7.5%)	+0.9	21 (9)	20 (9)

Largest class change: News / medical media -1.9pp. Among the top new domains, answers citing by type: consumer / retail (incl. pharmacy) 17; news / medical media 4; unclassified 3; regulator / government health body 2; market research 2; clinic / hospital / care provider 1.

Top new domains in the later cycle (public entities named; others typed with an opaque id): [consumer / retail (incl. pharmacy), UAE domain, H001] — 12 answers (10 naming); [consumer / retail (incl. pharmacy), non-regional domain, H002] — 2 answers (1 naming); emedicine.medscape.com — 2 answers (1 naming); ods.od.nih.gov — 2 answers (1 naming); [consumer / retail (incl. pharmacy), UAE domain, H003] — 2 answers (2 naming); [market research, non-regional domain, H004] — 2 answers (1 naming); biospace.com — 2 answers (2 naming); [clinic / hospital / care provider, non-regional domain, H005] — 1 answers (1 naming); [unclassified, UAE domain, H006] — 1 answers (0 naming); [unclassified, UAE domain, H007] — 1 answers (0 naming); [consumer / retail (incl. pharmacy), non-regional domain, H008] — 1 answers (1 naming); [unclassified, non-regional domain, H009] — 1 answers (1 naming).

Most-cited domains absent from the later cycle: [market research, non-regional domain, H010] — 4 answers; [clinic / hospital / care provider, non-regional domain, H011] — 2 answers; [consumer / retail (incl. pharmacy), UAE domain, H012] — 2 answers; [consumer / retail (incl. pharmacy), non-regional domain, H013] — 1 answers; [unclassified, non-regional domain, H014] — 1 answers; [consumer / retail (incl. pharmacy), non-regional domain, H015] — 1 answers; ovid.com — 1 answers; [clinic / hospital / care provider, non-regional domain, H016] — 1 answers.

Perplexity, UAE: 6 Oct against 7 Oct (after window)

Governed naming: 17/54 (31.5%) on 2026-10-06, 40/54 (74.1%) on 2026-10-07. Answers with at least one citation: 60/60 and 60/60. Distinct domains: 199 and 115 (46 new in the later cycle, 130 absent from it).

Class	Earlier: links (share)	Later: links (share)	Change (pp)	Earlier: answers citing (naming)	Later: answers citing (naming)
Regulator / government health body	98 (8.3%)	32 (5.3%)	-3.0	28 (11)	17 (15)
Journal / literature / trial registry	78 (6.6%)	18 (3.0%)	-3.6	23 (8)	12 (9)
Congress	0 (0.0%)	0 (0.0%)	0.0	0 (0)	0 (0)
Sponsor / manufacturer	0 (0.0%)	0 (0.0%)	0.0	0 (0)	0 (0)
News / medical media	207 (17.5%)	45 (7.5%)	-10.0	37 (20)	25 (20)
Clinic / hospital / care provider	301 (25.5%)	116 (19.3%)	-6.2	46 (20)	35 (33)
Consumer / retail (incl. pharmacy)	339 (28.7%)	68 (11.3%)	-17.4	38 (14)	34 (24)
Other (classified)	118 (10.0%)	302 (50.3%)	+40.3	45 (17)	56 (42)
Unclassified	40 (3.4%)	19 (3.2%)	-0.2	15 (8)	14 (11)

Largest class change: Other (classified) +40.3pp. Among the top new domains, answers citing by type: market research 63; consumer / retail (incl. pharmacy) 7; other (public) 6; journal / literature / trial registry 3; unclassified 2; clinic / hospital / care provider 2; directory or platform 2.

Top new domains in the later cycle (public entities named; others typed with an opaque id): [market research, non-regional domain, H017] — 25 answers (16 naming); [market research, non-regional domain, H018] — 14 answers (6 naming); [market research, non-regional domain, H019] — 11 answers (8 naming); [market research, non-regional domain, H020] — 7 answers (4 naming); [market research, non-regional domain, H021] — 6 answers (5 naming); drugs.com — 6 answers (6 naming); [consumer / retail (incl. pharmacy), UAE domain, H022] — 5 answers (3 naming); pubmed.ncbi.nlm.nih.gov — 3 answers (1 naming); [unclassified, non-regional domain, H023] — 2 answers (2 naming); [clinic / hospital / care provider, non-regional domain, H024] — 2 answers (2 naming); [directory or platform, non-regional domain, H025] — 2 answers (2 naming); [consumer / retail (incl. pharmacy), non-regional domain, H026] — 2 answers (2 naming).

Most-cited domains absent from the later cycle: [consumer / retail (incl. pharmacy), UAE domain, H027] — 21 answers; [consumer / retail (incl. pharmacy), non-regional domain, H028] — 19 answers; [consumer / retail (incl. pharmacy), non-regional domain, H029] — 13 answers; [clinic / hospital / care provider, UAE domain, H030] — 13 answers; zawya.com — 9 answers; [clinic / hospital / care provider, UAE domain, H031] — 9 answers; globenewswire.com — 8 answers; [clinic / hospital / care provider, UAE domain, H032] — 8 answers.

What this shows. The kinds of source each engine cited in each window, and what changed in the two cycle pairs where naming stepped. **What it does not show.** That a class change caused a naming change (both windows share questions, and the engine's own retrieval and model can change on the same day); that a cited page said what the class implies; or anything about engines' consumer products. Host-level classes are approximate: a host's class is its publisher's type, not the page's content.

11.5.2 E2 band-rule readings

Protocol §4: the band is the v4 single-cycle range "shifted by the §6 v4→v5 offset for that engine, with each end widened by the offset's 95% interval". The registered decision uses reading (i), I1 (D16). (ii) and (iii) are two alternative interpretations examined after the results; a shift-only band, which omits the registered widening, is shown for reference. The after-window rate is the registered pooled rate; bands use the registered limits and stored offset bounds, as the decision did. Signed distance: positive = beyond the nearer edge, negative = inside by that much.

Series	Engine	After K/N (rate)	Registered verdict	(i) I1 band → verdict (distance pp)	(ii) half-width → verdict (distance)	(iii) full width → verdict (distance)	Shift only (reference)	Flips under (ii)/(iii)
uae-glp1	Perplexity	243/378 (64.286%)	no movement	41.98%–64.36% → no movement (-0.074)	41.36%–63.74% → movement (+0.546)	36.42%–68.68% → no movement (-4.394)	46.30%–58.80% → movement (+5.486)	ii-shift-widen-half-width
uae-glp1	Gemini	206/378 (54.497%)	no movement	35.14%–78.42% → no movement (-19.357)	35.14%–78.42% → no movement (-19.357)	26.50%–87.06% → no movement (-27.997)	43.78%–69.78% → no movement (-10.717)	none
uae-glp1	ChatGPT	110/378 (29.101%)	no movement	14.23%–46.87% → no movement (-14.870)	14.23%–46.87% → no movement (-14.870)	9.91%–51.19% → no movement (-19.191)	18.55%–42.55% → no movement (-10.550)	none
uae-glp1	Claude	256/378 (67.725%)	movement	34.53%–67.33% → movement (+0.395)	34.84%–67.64% → movement (+0.085)	25.89%–76.59% → no movement (-8.865)	43.79%–58.69% → movement (+9.035)	iii-shift-widen-full-width
uae-glp1	Four engines	815/1512 (53.902%)	no movement	36.88%–53.98% → no movement (-0.078)	36.88%–53.98% → no movement (-0.078)	33.33%–57.53% → no movement (-3.628)	40.43%–50.43% → movement (+3.472)	none
saudi-t1d (post-reg.)	Perplexity	26/440 (5.909%)	no movement	3.77%–23.64% → no movement (-2.135)	3.47%–23.34% → no movement (-2.436)	2.56%–24.25% → no movement (-3.345)	4.38%–22.43% → no movement (-1.526)	none
saudi-t1d (post-reg.)	Gemini	40/440 (9.091%)	no movement	3.03%–30.30% → no movement (-6.064)	3.03%–30.30% → no movement (-6.059)	-2.42%–35.76% → no movement (-11.514)	8.49%–24.85% → no movement (-0.605)	none
saudi-t1d (post-reg.)	ChatGPT	81/440 (18.409%)	no movement	6.67%–35.15% → no movement (-11.740)	6.67%–35.15% → no movement (-11.740)	2.43%–39.39% → no movement (-15.980)	10.91%–30.91% → no movement (-7.500)	none
saudi-t1d (post-reg.)	Claude	28/437 (6.407%)	no movement	-1.21%–13.33% → no movement (-6.922)	-1.21%–13.33% → no movement (-6.927)	-3.02%–15.15% → no movement (-8.742)	0.61%–11.52% → no movement (-5.112)	none
saudi-t1d (post-reg.)	Four engines	175/1757 (9.960%)	no movement	6.49%–23.33% → no movement (-3.466)	6.57%–23.41% → no movement (-3.386)	4.68%–25.30% → no movement (-5.276)	8.46%–21.52% → no movement (-1.496)	none

What this shows. Which verdicts depend on the reading. uae-glp1 Perplexity (no movement under I1) flips under ii-shift-widen-half-width; uae-glp1 Claude (movement under I1) flips under iii-shift-widen-full-width. **What it does not show.** Which reading the author intended: I1 was fixed in D16 after per-cycle counts had been displayed. The registered verdicts are those of I1 and stand; this table never replaces them.

11.5.3 Cycle-level change estimate

mean after-window cycle rate minus mean v4 reference cycle rate minus the bridge point offset; cycles are the unit. A different estimand from the band rule. SE = sqrt(sd_after²/n_after + sd_before²/n_before + se_offset²); se_offset = (upper – lower)/3.92 from the bootstrap 95% interval (normal approximation); conservative variant uses max(offset – lower, upper – offset)/1.96. z and 95% CI by normal approximation; cycles treated as independent.

Series	Engine	Before: cycles, mean (SD)	After: cycles, mean (SD)	Offset (95% CI)	Change net of offset	SE (cycles / offset / total)	z	95% CI	z with conservative offset SE
uae-glp1	Perplexity	7, 52.1% (3.93pp)	7, 64.3% (19.03pp)	0.00 (-4.32 to +5.56)	+12.19pp	7.34 / 2.52 / 7.76	1.57	-3.0 to +27.4	1.55
uae-glp1	Gemini	13, 59.3% (7.49pp)	7, 54.5% (5.84pp)	-4.32 (-12.96 to +4.32)	-0.51pp	3.03 / 4.41 / 5.35	-0.10	-11.0 to +10.0	-0.10
uae-glp1	ChatGPT	13, 30.2% (5.97pp)	7, 29.1% (4.87pp)	-1.85 (-6.17 to +2.47)	+0.75pp	2.48 / 2.20 / 3.31	0.23	-5.7 to +7.2	0.23
uae-glp1	Claude	13, 47.9% (4.40pp)	7, 67.7% (3.36pp)	+3.09 (-6.17 to +11.73)	+16.77pp	1.76 / 4.57 / 4.89	3.43	+7.2 to +26.4	3.33
uae-glp1	Four engines	13, 45.6% (2.87pp)	7, 53.9% (4.72pp)	-0.77 (-4.32 to +2.78)	+9.10pp	1.95 / 1.81 / 2.66	3.42	+3.9 to +14.3	3.42
saudi-t1d (post-reg.)	Perplexity	11, 9.8% (5.59pp)	8, 5.9% (7.70pp)	+0.61 (0.00 to +1.82)	-4.49pp	3.20 / 0.46 / 3.24	-1.39	-10.8 to +1.9	-1.38
saudi-t1d (post-reg.)	Gemini	11, 9.0% (5.17pp)	8, 9.1% (2.92pp)	+4.85 (-0.61 to +10.30)	-4.71pp	1.87 / 2.78 / 3.35	-1.41	-11.3 to +1.9	-1.40
saudi-t1d (post-reg.)	ChatGPT	11, 17.4% (6.52pp)	8, 18.4% (3.82pp)	0.00 (-4.24 to +4.24)	+1.05pp	2.39 / 2.16 / 3.22	0.33	-5.3 to +7.4	0.33
saudi-t1d (post-reg.)	Claude	11, 5.0% (4.23pp)	8, 6.4% (4.50pp)	+0.61 (-1.21 to +2.42)	+0.86pp	2.04 / 0.93 / 2.24	0.38	-3.5 to +5.3	0.38
saudi-t1d (post-reg.)	Four engines	11, 10.3% (4.32pp)	8, 10.0% (2.73pp)	+1.52 (-0.45 to +3.33)	-1.83pp	1.62 / 0.96 / 1.89	-0.97	-5.5 to +1.9	-0.96

What this shows. Treating cycles as the unit, Claude's UAE change net of the offset is +16.8pp (95% CI +7.2 to +26.4, z = 3.43); Perplexity's is +12.2pp (-3.0 to +27.4, z = 1.57), its SE inflated by the after-window regime switch (SD 19.0pp). This is **a different estimand from the band rule**: it asks whether the mean level changed, not whether the pooled rate left a range widened by interval bounds. **What it does not show.** A registered verdict; attribution of any change to the congress; and it assumes independent cycles, a normal approximation, an additive offset on the percentage scale measured on one day, and a symmetric offset interval (the conservative column relaxes the last).

11.5.4 Per-cycle view against the bridged band

Each before-v5, congress-window and after-window cycle against the registered bridged band (I1). Wilson 95% intervals assume independent answers and are therefore too narrow (answers to one question are correlated). "Naive Wilson interval wholly outside band" marks a cycle whose whole Wilson interval lies outside the band. Single-cycle Wilson intervals are illustrative binomial intervals; the fixed prompt battery does not establish the independent Bernoulli sampling model they assume, and they are not used to validate the registered movement decision (§8.3). Descriptive; the registered rule pools the after window.

UAE, Wegovy (registered)

Engine	Bridged band	09-27 (before-v5)	09-29 (during)	10-04 (after)	10-06 (after)	10-07 (after)	10-08 (after)	10-09 (after)	10-10 (after)	10-11 (after)	After cycles outside (above / below)
Perplexity	41.98%–64.36%	25/52 48.1% [35.1%–61.3%] inside	22/54 40.7% [28.7%–54.0%] below +1.24	23/54 42.6% [30.3%–55.8%] inside	17/54 31.5% [20.7%–44.7%] below +10.50	40/54 74.1% [61.1%–83.9%] above +9.71	41/54 75.9% [63.0%–85.4%] above +11.57	40/54 74.1% [61.1%–83.9%] above +9.71	39/54 72.2% [59.1%–82.4%] above +7.86	43/54 79.6% [67.1%–88.2%] above +15.27 (naive Wilson interval wholly outside band)	6/7 (5 / 1)
Gemini	35.14%–78.42%	34/54 63.0% [49.6%–74.6%] inside	35/54 64.8% [51.5%–76.2%] inside	31/54 57.4% [44.2%–69.7%] inside	32/54 59.3% [46.0%–71.3%] inside	32/54 59.3% [46.0%–71.3%] inside	30/54 55.6% [42.4%–68.0%] inside	28/54 51.9% [38.9%–64.6%] inside	30/54 55.6% [42.4%–68.0%] inside	23/54 42.6% [30.3%–55.8%] inside	0/7 (0 / 0)
ChatGPT	14.23%–46.87%	15/54 27.8% [17.6%–40.9%] inside	15/54 27.8% [17.6%–40.9%] inside	13/54 24.1% [14.6%–37.0%] inside	16/54 29.6% [19.1%–42.8%] inside	16/54 29.6% [19.1%–42.8%] inside	14/54 25.9% [16.1%–38.9%] inside	16/54 29.6% [19.1%–42.8%] inside	21/54 38.9% [27.0%–52.2%] inside	14/54 25.9% [16.1%–38.9%] inside	0/7 (0 / 0)
Claude	34.53%–67.33%	22/54 40.7% [28.7%–54.0%] inside	31/54 57.4% [44.2%–69.7%] inside	36/54 66.7% [53.4%–77.8%] inside	37/54 68.5% [55.3%–79.3%] above +1.19	40/54 74.1% [61.1%–83.9%] above +6.74	37/54 68.5% [55.3%–79.3%] above +1.19	36/54 66.7% [53.4%–77.8%] inside	34/54 63.0% [49.6%–74.6%] inside	36/54 66.7% [53.4%–77.8%] inside	3/7 (3 / 0)
Four engines	36.88%–53.98%	96/214 44.9% [38.4%–51.6%] inside	103/216 47.7% [41.1%–54.3%] inside	103/216 47.7% [41.1%–54.3%] inside	102/216 47.2% [40.7%–53.9%] inside	128/216 59.3% [52.6%–65.6%] above +5.28	122/216 56.5% [49.8%–62.9%] above +2.50	120/216 55.6% [48.9%–62.0%] above +1.58	124/216 57.4% [50.7%–63.8%] above +3.43	116/216 53.7% [47.0%–60.2%] inside	4/7 (4 / 0)

Saudi Arabia, Tzield (post-registration, D10)

Engine	Bridged band	10-04 (after)	10-05 (after)	10-06 (after)	10-07 (after)	10-08 (after)	10-09 (after)	10-10 (after)	10-11 (after)	After cycles outside (above / below)
Perplexity	3.77%–23.64%	5/55 9.1% [4.0%–19.6%] inside	6/55 10.9% [5.1%–21.8%] inside	12/55 21.8% [13.0%–34.4%] inside	1/55 1.8% [0.3%–9.6%] below +1.95	0/55 0.0% [0.0%–6.5%] below +3.77	0/55 0.0% [0.0%–6.5%] below +3.77	2/55 3.6% [1.0%–12.3%] below +0.13	0/55 0.0% [0.0%–6.5%] below +3.77	5/8 (0 / 5)
Gemini	3.03%–30.30%	4/55 7.3% [2.9%–17.3%] inside	8/55 14.5% [7.6%–26.2%] inside	5/55 9.1% [4.0%–19.6%] inside	6/55 10.9% [5.1%–21.8%] inside	4/55 7.3% [2.9%–17.3%] inside	4/55 7.3% [2.9%–17.3%] inside	6/55 10.9% [5.1%–21.8%] inside	3/55 5.5% [1.9%–14.8%] inside	0/8 (0 / 0)
ChatGPT	6.67%–35.15%	9/55 16.4% [8.9%–28.3%] inside	6/55 10.9% [5.1%–21.8%] inside	12/55 21.8% [13.0%–34.4%] inside	9/55 16.4% [8.9%–28.3%] inside	10/55 18.2% [10.2%–30.3%] inside	11/55 20.0% [11.6%–32.4%] inside	12/55 21.8% [13.0%–34.4%] inside	12/55 21.8% [13.0%–34.4%] inside	0/8 (0 / 0)
Claude	-1.21%–13.33%	0/55 0.0% [0.0%–6.5%] inside	5/52 9.6% [4.2%–20.6%] inside	5/55 9.1% [4.0%–19.6%] inside	6/55 10.9% [5.1%–21.8%] inside	0/55 0.0% [0.0%–6.5%] inside	3/55 5.5% [1.9%–14.8%] inside	3/55 5.5% [1.9%–14.8%] inside	6/55 10.9% [5.1%–21.8%] inside	0/8 (0 / 0)
Four engines	6.49%–23.33%	18/220 8.2% [5.2%–12.6%] inside	25/217 11.5% [7.9%–16.4%] inside	34/220 15.5% [11.3%–20.8%] inside	22/220 10.0% [6.7%–14.7%] inside	14/220 6.4% [3.8%–10.4%] below +0.13	18/220 8.2% [5.2%–12.6%] inside	23/220 10.5% [7.1%–15.2%] inside	21/220 9.5% [6.3%–14.1%] inside	1/8 (0 / 1)

The 27 Sep UAE cycle (before-v5, admitted). Claude named Wegovy in 22/54 answers (40.7%), inside its bridged band, and in 31/54 (57.4%) on 29 Sep (congress window, inside). Claude's rise therefore began between the 27 Sep and 29 Sep cycles, both on the v5 instrument.

What this shows. Per cycle, Claude's after-window naming lies outside its bridged band in 3 of 7 cycles (3 above); Perplexity's in 6 of 7 (5 above, 1 below): its pooled "inside" verdict averages a low segment and a high one. **What it does not show.** A registered verdict (the rule is pooled), or why either engine moved.

11.5.5 Corrected bridge self-check

Expected range of n independent draws $\propto d2(n)$. With no between-day component, $E[R3]/E[Rk] = d2(3)/d2(k)$. d2-adjusted ratio = $(R3/d2(3)) / (Rk/d2(k))$: ≈ 1 when the same-day spread accounts for the band, < 1 when between-day variation exceeds it. SDs: between-cycle (k – 1 df) and within-day from the three same-day runs (2 df per arm; pooled v4+v5, 4 df). One-sided F test of between > within. Within-day SDs rest on 2 (or 4 pooled) degrees of freedom and are very imprecise. Singapore is at ceiling and is omitted. Band widths use the exact single-cycle range of the cycles that define each band (UAE Perplexity: the 7 cycles from 9 Sep).

Series	Engine	k	Band width (pp)	v4 runs	v5 runs	Same-day spread v4 / v5 (pp)	Registered ratio (v4 spread / width)	Expected with no day effect, d2(3)/d2(k)	d2-adjusted ratio: v4 / v5 / mean	SD between cycles (pp)	SD within day v4 / v5 / pooled (pp)	F (pooled), p one-sided	Material excess (ratio < 0.5): v4 runs / v4+v5 runs	F p < 0.05
uae-glp1	Perplexity	7	12.53	19/54, 16/54, 16/54	18/54, 15/54, 18/54	5.56 / 5.56	0.444	0.626	0.708 / 0.708 / 0.708	3.93	3.21 / 3.21 / 3.21	1.50, 0.362	no / no	no
uae-glp1	Gemini	13	25.93	31/54, 32/54, 30/54	31/54, 26/54, 29/54	3.70 / 9.26	0.143	0.507	0.281 / 0.704 / 0.493	7.49	1.85 / 4.66 / 3.55	4.46, 0.080	yes / yes	no
uae-glp1	ChatGPT	13	24.07	17/54, 16/54, 15/54	16/54, 15/54, 14/54	3.70 / 3.70	0.154	0.507	0.303 / 0.303 / 0.303	5.97	1.85 / 1.85 / 1.85	10.40, 0.018	yes / yes	yes
uae-glp1	Claude	13	14.81	37/54, 36/54, 33/54	36/54, 38/54, 37/54	7.41 / 3.70	0.500	0.507	0.985 / 0.493 / 0.739	4.40	3.85 / 1.85 / 3.02	2.12, 0.245	no / no	no
uae-glp1	Four engines	13	9.97	104/216, 100/216, 94/216	101/216, 94/216, 98/216	4.63 / 3.24	0.464	0.507	0.915 / 0.640 / 0.778	2.87	2.33 / 1.63 / 2.01	2.04, 0.257	no / no	no
saudi-t1d (post-reg.)	Perplexity	11	18.04	0/55, 1/55, 0/55	0/55, 1/55, 1/55	1.82 / 1.82	0.101	0.534	0.189 / 0.189 / 0.189	5.59	1.05 / 1.05 / 1.05	28.38, 0.003	yes / yes	yes
saudi-t1d (post-reg.)	Gemini	11	16.36	5/55, 3/55, 3/55	9/55, 2/54, 8/55	3.64 / 12.66	0.222	0.534	0.416 / 1.450 / 0.933	5.17	2.10 / 6.84 / 5.06	1.04, 0.530	yes / no	no
saudi-t1d (post-reg.)	ChatGPT	11	20.00	11/55, 9/55, 8/55	10/55, 7/55, 11/55	5.45 / 7.27	0.273	0.534	0.511 / 0.681 / 0.596	6.52	2.78 / 3.78 / 3.32	3.86, 0.102	no / no	no
saudi-t1d (post-reg.)	Claude	11	10.91	0/55, 1/55, 2/55	1/55, 2/55, 1/55	3.64 / 1.82	0.333	0.534	0.625 / 0.312 / 0.469	4.23	1.82 / 1.05 / 1.48	8.13, 0.029	no / yes	yes
saudi-t1d (post-reg.)	Four engines	11	13.06	16/220, 14/220, 13/220	20/220, 12/219, 21/220	1.36 / 4.07	0.104	0.534	0.196 / 0.584 / 0.390	4.32	0.69 / 2.23 / 1.65	6.86, 0.039	yes / yes	yes

Post hoc comparison of within-day and between-day variation (expected: Gemini, ChatGPT and Saudi Perplexity show excess between-day variation; UAE Claude and Perplexity do not), computed rather than assumed:

Series	Engine	Expected	Material on v4 runs (matches)	Material on v4+v5 runs (matches)	F test p < 0.05
uae-glp1	Gemini	excess	excess (yes)	excess (yes)	no
uae-glp1	ChatGPT	excess	excess (yes)	excess (yes)	yes
saudi-t1d	Perplexity	excess	excess (yes)	excess (yes)	yes
uae-glp1	Claude	no excess	no excess (yes)	no excess (yes)	no
uae-glp1	Perplexity	no excess	no excess (yes)	no excess (yes)	no

What this shows. "Material" excess is defined here as a between-day sigma at least twice the within-day sigma (d2-adjusted ratio < 0.5); the threshold was chosen after the ratios were seen, so the classification is descriptive, and the F test beside it needs no threshold. On the v4 runs, the registered comparison, it holds for UAE Gemini, UAE ChatGPT, Saudi Perplexity, Saudi Gemini; with the v5 runs included it holds for UAE Gemini, UAE ChatGPT, Saudi Perplexity, Saudi Claude; the F test (4 df within) reaches p < 0.05 for UAE ChatGPT, Saudi Perplexity, Saudi Claude. UAE Claude's v4 same-day spread is what the n-difference alone predicts (ratio 0.99), so for the engine carrying the declared movement the data show no excess between-day variation. The classification of the smaller engine-series moves with the run set, because three runs give a very imprecise within-day estimate. The registered check compared a 3-run range with a k-cycle range, which is not like with like; as registered it was ill-posed as well as failed. **What it does not show.** The day × setting interaction (one bridge day cannot identify it), or that the offset measured at the 5 Oct operating point transfers to the band period.

11.5.6 Brand-only naming beside the registered measure

The registered definition, exactly. E2's naming outcome is the production scorer's brandMentioned for the answer, with the scan's brand names: the brand, every alias configured on the scan input, and, where the scan's domain grants molecule credit, the molecule the domain resolves for the brand. The names used in every admitted cycle were:

Series	Brand	Registered names (brand or molecule)	Molecule credit	Brand-only names (this item)
UAE, Wegovy (registered)	Wegovy	Wegovy, semaglutide	molecule credit on (semaglutide)	Wegovy
Singapore, Jardiance (registered)	Jardiance	Jardiance, empagliflozin	molecule credit off (alias list only)	Jardiance
Saudi Arabia, Tzield (post-registration, D10)	Tzield	Tzield, tepilizumab	molecule credit off (alias list only)	Tzield

So the registered measure counts an answer that names only the molecule. Brand-only naming reruns the same scorer with the molecule removed. No bridge offset exists for brand-only naming (the bridge measured the registered measure), so v5 windows against v4 windows are descriptive and unbridged.

UAE, Wegovy (registered) — governed answers; registered (brand or molecule) and brand-only, K/N (rate)

Window	Perplexity	Gemini	ChatGPT	Claude	Four engines
before-v4	309/689 (44.8%) · brand-only 279 (40.5%)	416/701 (59.3%) · brand-only 392 (55.9%)	212/702 (30.2%) · brand-only 197 (28.1%)	336/702 (47.9%) · brand-only 319 (45.4%)	1273/2794 (45.6%) · brand-only 1187 (42.5%)
before-v5	25/52 (48.1%) · brand-only 21 (40.4%)	34/54 (63.0%) · brand-only 34 (63.0%)	15/54 (27.8%) · brand-only 12 (22.2%)	22/54 (40.7%) · brand-only 21 (38.9%)	96/214 (44.9%) · brand-only 88 (41.1%)
during	22/54 (40.7%) · brand-only 18 (33.3%)	35/54 (64.8%) · brand-only 31 (57.4%)	15/54 (27.8%) · brand-only 14 (25.9%)	31/54 (57.4%) · brand-only 29 (53.7%)	103/216 (47.7%) · brand-only 92 (42.6%)
after	243/378 (64.3%) · brand-only 196 (51.9%)	206/378 (54.5%) · brand-only 199 (52.6%)	110/378 (29.1%) · brand-only 103 (27.2%)	256/378 (67.7%) · brand-only 254 (67.2%)	815/1512 (53.9%) · brand-only 752 (49.7%)

Singapore, Jardiance (registered) — governed answers; registered (brand or molecule) and brand-only, K/N (rate)

Window	Perplexity	Gemini	ChatGPT	Claude	Four engines
before-v4	283/283 (100.0%) · brand-only 282 (99.6%)	281/284 (98.9%) · brand-only 281 (98.9%)	285/285 (100.0%) · brand-only 285 (100.0%)	284/285 (99.6%) · brand-only 281 (98.6%)	1133/1137 (99.8%) · brand-only 1129 (99.3%)
during	19/19 (100.0%) · brand-only 19 (100.0%)	19/19 (100.0%) · brand-only 19 (100.0%)	19/19 (100.0%) · brand-only 19 (100.0%)	19/19 (100.0%) · brand-only 19 (100.0%)	76/76 (100.0%) · brand-only 76 (100.0%)
after	133/133 (100.0%) · brand-only 133 (100.0%)	133/133 (100.0%) · brand-only 133 (100.0%)	133/133 (100.0%) · brand-only 133 (100.0%)	133/133 (100.0%) · brand-only 133 (100.0%)	532/532 (100.0%) · brand-only 532 (100.0%)

Saudi Arabia, Tzield (post-registration, D10) — governed answers; registered (brand or molecule) and brand-only, K/N (rate)

Window	Perplexity	Gemini	ChatGPT	Claude	Four engines
before-v4	58/589 (9.8%) · brand-only 27 (4.6%)	54/603 (9.0%) · brand-only 36 (6.0%)	105/605 (17.4%) · brand-only 105 (17.4%)	30/605 (5.0%) · brand-only 28 (4.6%)	247/2402 (10.3%) · brand-only 196 (8.2%)
after	26/440 (5.9%) · brand-only 19 (4.3%)	40/440 (9.1%) · brand-only 20 (4.5%)	81/440 (18.4%) · brand-only 81 (18.4%)	28/437 (6.4%) · brand-only 27 (6.2%)	175/1757 (10.0%) · brand-only 147 (8.4%)

What this shows. In the UAE, Claude named the brand itself in 45.4% of v4 answers and 67.2% of after-window answers (registered: 47.9% → 67.7%); Perplexity 40.5% → 51.9% (registered 44.8% → 64.3%; v4 here is all v4 cycles, not the post-step reference). Answers credited by the molecule alone are the gap between the two rates. Cross-checks: the scorer's own canonical-name flag disagrees with the brand-only rescoring in 0 governed answers, and 0 answers are brand-only-named but not registered-named. **What it does not show.** A bridged brand-only change, a brand-only band, or a verdict.

11.5.7 Search rate, and attribution among answers with at least one citation

What "searched" can mean, per API, in this instrument. The pull stores, per answer, the cited links and the search results the request path kept. In every answer the two are empty together or non-empty together, so "at least one citation" and "any searchResults" coincide here. Neither records a search the model ran without citing anything, so both are lower bounds on search invocation.

Engine	Endpoint as configured	What counts as a citation	What searchResults holds
ChatGPT	gpt-4o-mini, OpenAI Responses API with the web_search tool; the model decides per request whether to search	url_citation annotations on the final message, deduplicated by URL	the same cited URLs and titles (no snippets); uncited results are not returned
Claude	claude-haiku-4-5, Messages API with the web_search_20250305 server tool, max_uses 5; the model decides whether to search	web_search_result_location citations in text blocks, deduplicated by URL	the cited results only; retrieved-but-uncited results are read for dates and then discarded
Gemini	gemini-2.5-flash with Google Search grounding; grounding is model-decided	groundingChunks web URIs (Google-hosted redirect links, one per chunk)	the same chunks, with the source domain as the title
Perplexity	sonar, chat completions; retrieval on every request	the response's citations list (or search_results URLs)	the API's search_results

UAE, Wegovy (registered) — answers with at least one citation / captured answers (rate); mean citations and distinct domains per citing answer

Window	Perplexity	Gemini	ChatGPT	Claude
before-v4	765/765 (100.0%); 19.1 cit., 13.6 dom.	770/779 (98.8%); 27.2 cit., 23.6 dom.	521/780 (66.8%); 7.4 cit., 6.5 dom.	754/780 (96.7%); 8.1 cit., 7.4 dom.
before-v5	58/58 (100.0%); 19.4 cit., 12.3 dom.	60/60 (100.0%); 27.6 cit., 23.5 dom.	39/60 (65.0%); 7.6 cit., 6.6 dom.	57/60 (95.0%); 8.2 cit., 7.5 dom.
during	60/60 (100.0%); 19.8 cit., 12.2 dom.	60/60 (100.0%); 25.9 cit., 22.3 dom.	40/60 (66.7%); 7.5 cit., 6.9 dom.	57/60 (95.0%); 8.4 cit., 7.8 dom.
after	420/420 (100.0%); 12.8 cit., 9.9 dom.	418/420 (99.5%); 26.9 cit., 22.9 dom.	278/420 (66.2%); 7.5 cit., 6.5 dom.	405/420 (96.4%); 8.0 cit., 7.3 dom.

Singapore, Jardiance (registered) — answers with at least one citation / captured answers (rate); mean citations and distinct domains per citing answer

Window	Perplexity	Gemini	ChatGPT	Claude
before-v4	385/385 (100.0%); 18.2 cit., 13.2 dom.	380/389 (97.7%); 16.4 cit., 12.0 dom.	390/390 (100.0%); 4.1 cit., 3.2 dom.	390/390 (100.0%); 5.2 cit., 4.6 dom.
during	26/26 (100.0%); 19.8 cit., 13.2 dom.	26/26 (100.0%); 18.5 cit., 13.3 dom.	26/26 (100.0%); 4.7 cit., 3.6 dom.	26/26 (100.0%); 4.5 cit., 4.1 dom.
after	182/182 (100.0%); 12.2 cit., 8.4 dom.	181/182 (99.5%); 17.6 cit., 12.4 dom.	182/182 (100.0%); 4.0 cit., 3.2 dom.	182/182 (100.0%); 4.4 cit., 4.1 dom.

Saudi Arabia, Tzield (post-registration, D10) — answers with at least one citation / captured answers (rate); mean citations and distinct domains per citing answer

Window	Perplexity	Gemini	ChatGPT	Claude
before-v4	600/600 (100.0%); 19.1 cit., 14.2 dom.	589/614 (95.9%); 23.7 cit., 19.0 dom.	372/616 (60.4%); 4.8 cit., 4.2 dom.	598/616 (97.1%); 7.4 cit., 6.6 dom.
after	448/448 (100.0%); 13.4 cit., 10.8 dom.	436/448 (97.3%); 23.2 cit., 18.6 dom.	261/448 (58.3%); 5.0 cit., 4.6 dom.	434/445 (97.5%); 7.2 cit., 6.4 dom.

UAE, per cycle: governed answers that cited / governed answers, and naming among citing and non-citing governed answers

Cycle date	Window	Perplexity	Gemini	ChatGPT	Claude
2026-08-22	before-v4	49/49; named 19/49 citing, 0/0 not	52/54; named 36/52 citing, 1/2 not	33/54; named 19/33 citing, 0/21 not	53/54; named 23/53 citing, 0/1 not
2026-08-24	before-v4	54/54; named 21/54 citing, 0/0 not	54/54; named 30/54 citing, 0/0 not	34/54; named 14/34 citing, 0/20 not	53/54; named 28/53 citing, 0/1 not
2026-08-26	before-v4	54/54; named 19/54 citing, 0/0 not	54/54; named 29/54 citing, 0/0 not	35/54; named 18/35 citing, 0/19 not	50/54; named 26/50 citing, 0/4 not
2026-08-28	before-v4	54/54; named 18/54 citing, 0/0 not	54/54; named 29/54 citing, 0/0 not	35/54; named 24/35 citing, 0/19 not	53/54; named 22/53 citing, 0/1 not
2026-08-30	before-v4	53/53; named 20/53 citing, 0/0 not	52/54; named 33/52 citing, 1/2 not	34/54; named 17/34 citing, 0/20 not	52/54; named 28/52 citing, 0/2 not
2026-09-04	before-v4	54/54; named 19/54 citing, 0/0 not	53/54; named 26/53 citing, 0/1 not	33/54; named 16/33 citing, 0/21 not	52/54; named 28/52 citing, 0/2 not
2026-09-09	before-v4	54/54; named 28/54 citing, 0/0 not	53/54; named 31/53 citing, 0/1 not	33/54; named 14/33 citing, 0/21 not	53/54; named 26/53 citing, 0/1 not

Cycle date	Window	Perplexity	Gemini	ChatGPT	Claude
2026-09-11	before-v4	54/54; named 27/54 citing, 0/0 not	53/54; named 34/53 citing, 0/1 not	35/54; named 17/35 citing, 0/19 not	51/54; named 30/51 citing, 0/3 not
2026-09-14	before-v4	51/51; named 28/51 citing, 0/0 not	53/53; named 26/53 citing, 0/0 not	35/54; named 18/35 citing, 0/19 not	52/54; named 27/52 citing, 0/2 not
2026-09-16	before-v4	51/51; named 30/51 citing, 0/0 not	54/54; named 40/54 citing, 0/0 not	34/54; named 16/34 citing, 0/20 not	51/54; named 23/51 citing, 0/3 not
2026-09-21	before-v4	53/53; named 27/53 citing, 0/0 not	54/54; named 33/54 citing, 0/0 not	33/54; named 11/33 citing, 0/21 not	52/54; named 25/52 citing, 0/2 not
2026-09-23	before-v4	54/54; named 25/54 citing, 0/0 not	53/54; named 32/53 citing, 0/1 not	34/54; named 13/34 citing, 0/20 not	52/54; named 26/52 citing, 0/2 not
2026-09-25	before-v4	54/54; named 28/54 citing, 0/0 not	53/54; named 35/53 citing, 0/1 not	35/54; named 15/35 citing, 0/19 not	52/54; named 24/52 citing, 0/2 not
2026-09-27	before-v5	52/52; named 25/52 citing, 0/0 not	54/54; named 34/54 citing, 0/0 not	33/54; named 15/33 citing, 0/21 not	51/54; named 22/51 citing, 0/3 not
2026-09-29	during	54/54; named 22/54 citing, 0/0 not	54/54; named 35/54 citing, 0/0 not	34/54; named 15/34 citing, 0/20 not	51/54; named 31/51 citing, 0/3 not
2026-10-04	after	54/54; named 23/54 citing, 0/0 not	54/54; named 31/54 citing, 0/0 not	34/54; named 13/34 citing, 0/20 not	51/54; named 36/51 citing, 0/3 not
2026-10-06	after	54/54; named 17/54 citing, 0/0 not	54/54; named 32/54 citing, 0/0 not	34/54; named 16/34 citing, 0/20 not	52/54; named 37/52 citing, 0/2 not
2026-10-07	after	54/54; named 40/54 citing, 0/0 not	54/54; named 32/54 citing, 0/0 not	34/54; named 16/34 citing, 0/20 not	52/54; named 40/52 citing, 0/2 not
2026-10-08	after	54/54; named 41/54 citing, 0/0 not	54/54; named 30/54 citing, 0/0 not	35/54; named 14/35 citing, 0/19 not	52/54; named 37/52 citing, 0/2 not
2026-10-09	after	54/54; named 40/54 citing, 0/0 not	53/54; named 28/53 citing, 0/1 not	32/54; named 16/32 citing, 0/22 not	52/54; named 36/52 citing, 0/2 not
2026-10-10	after	54/54; named 39/54 citing, 0/0 not	54/54; named 30/54 citing, 0/0 not	33/54; named 21/33 citing, 0/21 not	52/54; named 34/52 citing, 0/2 not
2026-10-11	after	54/54; named 43/54 citing, 0/0 not	53/54; named 23/53 citing, 0/1 not	34/54; named 14/34 citing, 0/20 not	52/54; named 36/52 citing, 0/2 not

E2 attribution reading (I3) among governed after-window answers that cited at least one source

Series	Engine	All governed: qualifying (named) / others (named), p	Citing only: qualifying (named) / others (named), p	Min. attainable p (citing only)	Non-citing answers (named)
uae-glp1	Perplexity	1 (1) / 377 (242), 1	1 (1) / 377 (242), 1	0.3571	0 (0)
uae-glp1	Gemini	1 (1) / 377 (205), 1	1 (1) / 375 (205), 1	0.4521	2 (0)
uae-glp1	ChatGPT	0 (0) / 378 (110), —	0 (0) / 236 (110), —	—	142 (0)
uae-glp1	Claude	0 (0) / 378 (256), —	0 (0) / 363 (256), —	—	15 (0)
saudi-t1d	Perplexity	3 (1) / 437 (25), 0.1674	3 (1) / 437 (25), 0.1674	0.0002	0 (0)
saudi-t1d	Gemini	1 (0) / 439 (40), 1	1 (0) / 427 (40), 1	0.0935	12 (0)
saudi-t1d	ChatGPT	0 (0) / 440 (81), —	0 (0) / 253 (78), —	—	187 (3)
saudi-t1d	Claude	0 (0) / 437 (28), —	0 (0) / 426 (28), —	—	11 (0)

What this shows. How often each endpoint returned cited sources: in the UAE after window Claude 96.4%, Perplexity 100.0%, Gemini 99.5%, ChatGPT 66.2%. Claude's citation-presence rate was unchanged at 51/54 scored answers in each of the two cycles (27 Sep and 29 Sep). The study did not record issued queries or uncited retrieval results, so it cannot establish unchanged search behaviour. Restricting attribution to answers that cited leaves Claude with no qualifying answer, so its reading is unchanged (movement without verified congress sources), and leaves Perplexity's single qualifying answer unable to reach $p < 0.05$. **What it does not show.** Search invocation itself (not stored), the queries issued, or retrieved-but-uncited results; a citation count compares different objects across APIs.

11.5.8 Attribution sensitivity: unverified answers excluded, not counted as negative

Registered I3: Fisher exact, two-sided, after-window governed answers of the engine, E1-qualifying against all others. Sensitivity: others that cite an unverified (unreadable) new page are removed from the comparison group. Minimum attainable p: the smallest two-sided Fisher p any split of the qualifying answers could give with both margins fixed.

Series	Engine	Registered verdict	Registered: qualifying (named) / others (named), p	Min. attainable p (registered; frozen E1 classification)	Qualifying needed for $p < 0.05$	Others citing an unverified page, removed (named)	Sensitivity: qualifying (named) / others (named), p	Min. attainable p (sensitivity)	Focused questions included: qualifying (named) / others (named), p
uae-glp1	Perplexity	congress sources without movement	1 (1) / 377 (242), 1	0.3571	3	76 (53)	1 (1) / 301 (189), 1	0.3709	5 (5) / 415 (280), 0.181
uae-glp1	Gemini	congress sources without movement	1 (1) / 377 (205), 1	0.455	4	282 (157)	1 (1) / 95 (48), 1	0.4896	1 (1) / 419 (247), 1
uae-glp1	ChatGPT	neither	0 (0) / 378 (110), —	—	3	43 (12)	0 (0) / 335 (98), —	—	0 (0) / 420 (152), —
uae-glp1	Claude	movement without congress sources	0 (0) / 378 (256), —	—	3	76 (46)	0 (0) / 302 (210), —	—	0 (0) / 420 (298), —
saudi-t1d	Perplexity	congress sources without movement	3 (1) / 437 (25), 0.1674	0.0002	2	63 (4)	3 (1) / 374 (21), 0.1655	0.0002	3 (1) / 445 (33), 0.2113
saudi-t1d	Gemini	congress sources without movement	1 (0) / 439 (40), 1	0.0909	2	407 (40)	1 (0) / 32 (0), 1	1	1 (0) / 447 (48), 1
saudi-t1d	ChatGPT	neither	0 (0) / 440 (81), —	—	2	32 (7)	0 (0) / 408 (74), —	—	0 (0) / 448 (89), —
saudi-t1d	Claude	neither	0 (0) / 437 (28), —	—	2	72 (2)	0 (0) / 365 (26), —	—	1 (1) / 444 (35), 0.0809

E1 qualifying answers by question class (all captured answers; focused questions name the brand in the question and are excluded from the E2 governed population, so their qualifying answers never enter the registered attribution test)

Series	Window	Engine	Qualifying answers	Focused	Brand-neutral (governed)	Captured answers (focused)
uae-glp1	after	Perplexity	5	4	1	420 (42)
uae-glp1	after	Gemini	1	0	1	420 (42)
saudi-t1d	after	Perplexity	3	0	3	448 (8)
saudi-t1d	after	Gemini	1	0	1	448 (8)
saudi-t1d	after	Claude	1	1	0	445 (8)

What this shows. Under the frozen E1 classification, no split of the qualifying answers could have reached $p < 0.05$ for UAE Perplexity (1 qualifying, minimum p 0.3571); UAE Gemini (1 qualifying, minimum p 0.455); Saudi Gemini (1 qualifying, minimum p 0.0909): for them the registered test could not have returned "congress-linked" whatever those answers said. Only Saudi Perplexity (3 qualifying, minimum p 0.0002, observed p 0.1674) had enough qualifying answers for the test to be informative. UAE ChatGPT, UAE Claude, Saudi ChatGPT, Saudi Claude had no qualifying answer, so their reading rests on E1 = 0 alone. Removing other answers that cite an unverified page brings no test below $p < 0.05$ and changes no reading. Of 11 E1-qualifying answers in the congress and after windows, 5 answered focused questions and so sit outside the test. For UAE Gemini under the corrected classification, see §10.2 (no verdict changes). **What it does not show.** That the congress did not stand behind a movement: "movement without congress sources" means movement without *verified* congress sources in a test that was structurally unable to detect a link.

11.5.9 Saudi E3-i: type 1 use-context recode (model-coded sensitivity)

Model-coded, by the same model family as every other coder in the study (Claude); not human coding and not validation. Codes: docs/evidence/easd-2026-saudi-e3i-use-context.json, one per sampled E3-i naming coded as a type 1 use (stated, absent or wrong), keyed by answer sha256 and agent, each with its span and a note. Rules (from the file):

- R1 glycaemic: the answer presents the type 1 use as an adjunct or add-on to insulin, or states a glucose-related purpose or outcome for it (glycaemic control, blood sugar, HbA1c, time in range, insulin-dose or insulin-requirement reduction). A use described only as an 'adjunct' or 'add-on' in type 1 counts as insulin-adjunct use. Weight named alongside a glucose purpose does not change the code; it sets weightAlsoStated.
- R2 weight-management: the type 1 use is presented for obesity, overweight or weight loss in a person with type 1, and no glucose-related purpose and no adjunct-to-insulin framing is stated.
- R3 unspecified: neither R1 nor R2 — the molecule is named for type 1 without a stated purpose (product lists, 'may be prescribed off-label', 'specialist uses in selected circumstances', 'potential benefits in type 1 management'), or only for cardiovascular or kidney outcomes, or for unnamed comorbidities.
- R4 precedence: R1 is tested first, then R2, then R3. Context is read from the coded span and its paragraph only; status language (approved, off-label) never decides the context.
- R5 flags, for re-cutting: weightAlsoStated = weight loss or weight management is named as a purpose or outcome of the same use; populationOverweightOrObese = the use or the cited study is described in people with type 1 and overweight or obesity.
- R6 spans: the frozen coded span, with links and URLs replaced by '[link removed]' so no cited site is named in this file.

Window	Glycaemic / weight-management / unspecified	All type 1 uses: no status statement / namings (share, clustered 95%)	Glycaemic only	Glycaemic, no weight framing	Glycaemic, not an overweight population	Unspecified only
before-v4	44 / 0 / 12	38/56 (67.9%, 44.4%–87.3%)	30/44 (68.2%, 46.0%–87.0%)	7/10 (70.0%, 36.4%–100.0%)	20/31 (64.5%, 45.2%–86.7%)	8/12 (66.7%, 25.0%–100.0%)
after	18 / 0 / 21	25/39 (64.1%, 40.6%–82.0%)	16/18 (88.9%, 73.3%–100.0%)	8/8 (100.0%, 100.0%–100.0%)	12/14 (85.7%, 66.7%–100.0%)	9/21 (42.9%, 13.6%–75.0%)

What this shows. No sampled type 1 use was framed for weight management alone. Restricted to glycaemic (insulin-adjunct or glucose-control) use, the share of sampled glycaemic-use namings lacking qualifying indication-status wording under the study's rules (Saudi product-specific labeling not independently established) was 30/44 (68.2%) before and 16/18 (88.9%) after; the registered exploratory figure over all type 1 uses is 67.9% and 64.1%. The small fall in the registered figure is a change of mix: the after window holds more namings with no stated purpose (21 of 39, against 12 of 56 before), and those carry a status statement more often (no statement 42.9% after). Within glycaemic use the share did not fall. The clustered intervals for glycaemic use overlap (46.0%–87.0% against 73.3%–100.0%) and rest on 9 question slots after the congress. The 8/8 subgroup produces a degenerate empirical bootstrap interval because every observed item is positive. It should be read as a small-sample descriptive result, not evidence of zero uncertainty about an underlying rate. **What it does not show.** A congress effect, clinical appropriateness of any use, or anything independent of the model family that coded the E3-i items in the first place.

11.5.10 Per-answer outcome table

docs/evidence/computed/easd-2026-per-answer-outcomes.csv: 11883 rows, one per captured answer in the 64 admitted cycles. Columns: series, cycleSlug, cycleDate, window, engine, promptId, focused, inGovernedPopulation, namedRegistered, namedBrandOnly, searched, citationCount, e1Qualifying, citesUnverifiedNewPage, citesLocalPageP61, echoesLocalPageSentence. No question text and no answer text; promptId is the raw slot id (the evidence package maps it to an opaque id). The first line is the D24 label (a # comment line). Empty E1 fields are not applicable (E1 is defined for the congress and after windows only). **Check:** summing namedRegistered OVER inGovernedPopulation rows reproduces every governed per-cycle K, and the row count every N, in all 256 cycle × engine cells (0 mismatches; the script stops otherwise).

citesLocalPageP61 is true when the answer cites the item 12 page (matched by the same hash as item 12); echoesLocalPageSentence is true when the answer repeats S1 or S2 under item 12's 40-character rule. Neither the column names nor the values identify the page. **Item 12 gate:** from the table's own flags, Claude UAE governed naming reproduces item 12's split in all 4 windows and 22 UAE cycles (after window: 143/152 citing, 113/226 not citing; 29 Sep: 10/12 and 21/42), and from 29 Sep 152/164 citing answers and 0/316 non-citing answers echo a sentence (the script stops otherwise).

The e1Qualifying field records the frozen classification. The corrected classification additionally treats the documented 7 October UAE Gemini answer as qualifying; the public correction record robustness/easd-2026-e1-corrections.csv maps that answer to the table's public identifiers (cycle UAE-C19, question slot uae-q06). Frozen and corrected attribution calculations must use their respective classifications.

What this shows. Enough unit-level data to recompute E2, readings I1 and I2, I3 under the frozen classification (the corrected I3 also needs the public correction record), the clustered and per-cycle analyses, items 2–8 and item 12's central split here. **What it does not show.** Question wording, answer text, or which source an answer cited.

11.5.11 E1 corrections

11.5.11(a) Latency as a sampled detection bracket

OCTANE (LBA 59) was presented on 30 Sep 2026 at 09:45 UTC. The UAE series had no admitted cycle between 29 Sep (14:04 UTC) and 4 Oct; the 5 Oct cycle was excluded (D14). Each observed latency defines a sampled detection bracket (last negative sampled cycle, first positive sampled cycle).

Series	Engine	Readout	First qualifying cycle (day)	Previous admitted cycle (day)	Sampled detection bracket (days)	Excluded cycles in between
uae-glpl	Perplexity	OCTANE (switch to oral semaglutide, real-world), LBA 59	2026-10-04 14:05 UTC (4.2)	2026-09-29 14:04 UTC (-0.8), before presentation	[0, 4.2]	none

Series	Engine	Readout	First qualifying cycle (day)	Previous admitted cycle (day)	Sampled detection bracket (days)	Excluded cycles in between
uae- glp1	Gemini	OCTANE (switch to oral semaglutide, real-world), LBA 59	2026-10-06 14:05 UTC (6.2)	2026-10-04 14:05 UTC (4.2)	(4.2, 6.2]	2026-10-05 (UAE-C17, excluded)
saudi- t1d	Claude	EASD 2026 type 1 diabetes programme (agent set + teplizumab abstracts)	2026-10-08 16:34 UTC (9.4)	2026-10-07 16:34 UTC (8.4)	(8.4, 9.4]	none
uae- glp1	Perplexity	Other EASD 2026 data (not in the readout set)	2026-10-04 14:05 UTC (6.3)	2026-09-29 14:04 UTC (1.3)	(1.3, 6.3]	none
uae- glp1	Gemini	Other EASD 2026 data (not in the readout set)	2026-10-06 14:05 UTC (8.3)	2026-10-04 14:05 UTC (6.3)	(6.3, 8.3]	2026-10-05 (UAE-C17, excluded)
saudi- t1d	Perplexity	Other EASD 2026 data (not in the readout set)	2026-10-05 16:35 UTC (7.4)	2026-10-04 16:34 UTC (6.4)	(6.4, 7.4]	none
saudi- t1d	Gemini	Other EASD 2026 data (not in the readout set)	2026-10-07 16:34 UTC (9.4)	2026-10-06 16:34 UTC (8.4)	(8.4, 9.4]	none
saudi- t1d	Claude	Other EASD 2026 data (not in the readout set)	2026-10-08 16:34 UTC (10.4)	2026-10-07 16:34 UTC (9.4)	(9.4, 10.4]	none

What this shows. The last negative and first positive sampled cycles define observation brackets of (0, 4.2] days for Perplexity and (4.2, 6.2] days for Gemini. These brackets do not establish the engines' true evidence-uptake times, and the observation schedule and imperfect detection do not permit ranking the engines by uptake speed. **What it does not show.** Engine indexing speed, or publication-to-citation delay.

11.5.11(b) The text-route miss: answers stating the OCTANE result

E1 qualifying status in the frozen output. The corrected classification is in §10.2 and in the public correction record (§11.5.10).

Answer (sha256 prefix)	Cycle date	Engine	Figure	D4 hand check	Confirmed in the pull	Answer names OCTANE	E1 qualifying status in the frozen output (route)	Text ledger entry
e2c0f5d7d4e0	2026-10-06	Gemini	[Gemini text not reproduced, §13.3] (C-OCTANE-pct)	confirmed	yes	no	OCTANE (switch to oral semaglutide, real-world), LBA 59; Other EASD 2026 data (not in the readout set) (source)	none
70aea090029	2026-10-07	Gemini	[Gemini text not reproduced, §13.3] (C-OCTANE-pct)	confirmed	yes	no	no	none

What this shows. 2 Gemini answers state the OCTANE result (matching reference value C-OCTANE-pct; wording not reproduced, §13.3); the text route flagged neither, because neither names the trial or the congress. 1 reached E1 through its cited source; 1 is not counted by E1 at all. This is a **text-route recall limitation**: "0 answers reported a readout result" holds for answers that name a readout, not for answers that state one without naming it. Gemini's OCTANE latency (6 Oct cycle) is unchanged, because the first such answer is in the same cycle. **What it does not show.** The text route's recall for other readouts: the D4 check covered efficacy figures for the UAE medicine only.

11.5.11(c) Singapore E3 (registered secondary): investigational-agent namings by window

Window	Answers searched	Coded namings (correct / absent / wrong)	Agent matches on an independent re-search
before-v4	1554	0 (0 / 0 / 0)	0
before-v5	0	0 (0 / 0 / 0)	0
during	104	0 (0 / 0 / 0)	0
after	728	0 (0 / 0 / 0)	0

What this shows. The registered secondary result is confirmed from the final coding: no Singapore answer in any window named an in-scope investigational agent of the readout set (2386 answers searched; 0 out-of-scope matches). **What it does not show.** Anything about agents outside the readout set.

11.5.12 The page behind Claude's UAE rise (page P61)

Post-registration, descriptive and post hoc. Everything below is an association between a cited page and naming. It is not a causal estimate, and it never replaces a registered verdict.

11.5.12(a) What the page is

Type: **a local commercial blog post (UAE), August 2026, before the congress.** Typed from the page's own content, fetched read-only by the study coordinator on 11 Oct 2026 at 23:46 Dubai time (19:46 UTC). The page's date (August 2026, given to the month only) and the two sentences tested below come from that fetch. A fetch on 11 Oct cannot by itself establish what the page said when Claude cited it. Shared wording between a captured answer and a later-fetched page corroborates textual overlap. Without a contemporaneous archive, it does not establish the page's complete historical content, the direction of copying or exclusive source provenance. The local-business rule applies: its domain, URL and company are never written. It first appeared in Claude's resolved study citations in this series on 2026-09-29 (during), among the resolved citations examined; pre-congress Gemini destinations were unresolved.

Series	Engine	Window	Answers citing / captured answers	Naming among citing
uae- glp1	Claude	during	12/60	10/12
uae- glp1	Claude	after	152/420	143/152

Gemini cites through redirect links, resolved only for the congress and after windows; a Gemini citation of the page before 28 Sep could not be seen.

11.5.12(b) Textual overlap, by echo

An answer echoes a sentence when its normalised text shares a contiguous run of at least 40 characters with the sentence (exact or near-exact after normalisation: lowercase, every run of characters other than a–z, 0–9 and % made one space). Runs of at least 80 characters and of nearly the whole sentence (within 5 characters) are shown beside it. The sentences are held as hashes of 20-character windows and described in paraphrase only.

- **S1** (219 normalised characters): a class statement.
- **S2** (79 normalised characters): an efficacy-figure sentence.

Claude, UAE, every admitted cycle (captured answers):

Cycle date	Window	Answers	Citing the page	S1 echo ≥ 40: citing / not citing	S2 echo ≥ 40: citing / not	Either ≥ 80: citing / not	Either nearly whole sentence: citing / not	Longest run in a non-citing answer
2026-08-22	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-08-24	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	37
2026-08-26	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-08-28	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-08-30	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-04	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	35
2026-09-09	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-11	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-14	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-16	before-v4	60	0	0 / 1	0 / 0	0 / 0	0 / 0	46
2026-09-21	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	35
2026-09-23	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	35
2026-09-25	before-v4	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-27	before-v5	60	0	0 / 0	0 / 0	0 / 0	0 / 0	34
2026-09-29	during	60	12	10 / 0	0 / 0	9 / 0	7 / 0	34
2026-10-04	after	60	22	20 / 0	3 / 0	20 / 0	18 / 0	24
2026-10-06	after	60	23	23 / 0	3 / 0	21 / 0	22 / 0	30
2026-10-07	after	60	24	23 / 0	6 / 0	23 / 0	22 / 0	27
2026-10-08	after	60	21	21 / 0	7 / 0	20 / 0	18 / 0	34
2026-10-09	after	60	20	18 / 0	4 / 0	18 / 0	16 / 0	34
2026-10-10	after	60	23	20 / 0	5 / 0	20 / 0	18 / 0	34
2026-10-11	after	60	19	17 / 0	7 / 0	17 / 0	17 / 0	30

Echoes before 29 Sep (any engine, UAE): 2026-09-16 Claude, S1: a 46-character run starting at character 0 of 219, not citing the page.

All engines, UAE: answers echoing either sentence (≥ 40 / ≥ 80 characters; echoing and citing the page):

Window	Perplexity	Gemini	ChatGPT	Claude
before-v4	0 / 0 (0)	0 / 0 (0)	0 / 0 (0)	1 / 0 (0)
before-v5	0 / 0 (0)	0 / 0 (0)	0 / 0 (0)	0 / 0 (0)
during	0 / 0 (0)	0 / 0 (0)	0 / 0 (0)	10 / 9 (10)
after	0 / 0 (0)	0 / 0 (0)	0 / 0 (0)	142 / 139 (142)

What this shows. From 29 Sep, Claude's answers repeat the page's own wording: 152 of 164 answers that cite the page echo at least one of the two sentences, and 0 answers that do not cite it do. Before 29 Sep there is one echo at the 40-character threshold (2026-09-16, S1, 46 characters from character 0, not citing the page); none reaches 80 characters: a short run at the sentence's opening in common wording, not the page's distinctive wording. This corroborates textual overlap between the answers and these two sentences of the later-fetched page, read the same way as the diaTribe Tzield item (§7.1.7); it does not establish the page's complete historical content, the direction of copying or exclusive source provenance. **What it does not show.** The rest of the page at citation; or that the page caused anything.

11.5.12(c) Where Claude's UAE rise sits

Governed Claude UAE answers, naming K/N (rate), in answers that cite the page and answers that do not:

Window	Citing the page	Not citing
before-v4	—	336/702 (47.9%)
before-v5	—	22/54 (40.7%)
during	10/12 (83.3%)	21/42 (50.0%)
after	143/152 (94.1%)	113/226 (50.0%)

Cycle date	Window	Citing the page	Not citing
2026-09-27	before-v5	—	22/54 (40.7%)
2026-09-29	during	10/12 (83.3%)	21/42 (50.0%)
2026-10-04	after	20/22 (90.9%)	16/32 (50.0%)
2026-10-06	after	23/23 (100.0%)	14/31 (45.2%)
2026-10-07	after	24/24 (100.0%)	16/30 (53.3%)
2026-10-08	after	21/21 (100.0%)	16/33 (48.5%)

Cycle date	Window	Citing the page	Not citing
2026-10-09	after	18/20 (90.0%)	18/34 (52.9%)
2026-10-10	after	20/23 (87.0%)	14/31 (45.2%)
2026-10-11	after	17/19 (89.5%)	19/35 (54.3%)

By question. Each question's naming before the congress (13 v4 cycles and 27 Sep) against after, split by whether the page was cited in the after window. Questions are keyed by their text and shown as opaque ids; per-question rows are in the JSON. The groups hold questions asked in the after window; 3 question texts occur only before it and 0 only in it (the battery matches its first cycle at 57–60 of 60).

Questions	Count	Before: K/N (rate)	After: K/N (rate)	After, citing the page	After, not citing	Questions with before rate ≥ 50%
Drew the page in the after window	31	167/414 (40.3%)	171/217 (78.8%)	143/152 (94.1%)	28/65 (43.1%)	9
Never drew the page	23	181/312 (58.0%)	85/161 (52.8%)	—	85/161 (52.8%)	13

Question	Before K/N	After, citing the page	After, not citing
Q01	3/14 (21.4%)	6/6	0/1
Q02	2/14 (14.3%)	3/3	2/4
Q04	7/14 (50.0%)	7/7	0/0
Q06	3/14 (21.4%)	5/5	0/2
Q07	0/4 (0.0%)	0/3	0/4
Q11	5/14 (35.7%)	2/2	0/5
Q13	1/4 (25.0%)	1/1	0/6
Q15	11/14 (78.6%)	4/4	2/3
Q16	4/14 (28.6%)	7/7	0/0
Q19	6/14 (42.9%)	7/7	0/0
Q20	9/14 (64.3%)	7/7	0/0
Q22	4/14 (28.6%)	7/7	0/0
Q25	5/14 (35.7%)	1/1	2/6
Q27	4/14 (28.6%)	5/5	2/2
Q30	6/14 (42.9%)	5/5	1/2
Q31	1/14 (7.1%)	4/7	0/0
Q32	5/14 (35.7%)	7/7	0/0
Q33	1/14 (7.1%)	2/2	5/5
Q34	6/14 (42.9%)	1/1	0/6
Q36	4/14 (28.6%)	7/7	0/0
Q37	3/14 (21.4%)	6/6	0/1
Q38	8/14 (57.1%)	4/4	3/3
Q39	13/14 (92.9%)	1/1	6/6
Q40	9/14 (64.3%)	4/7	0/0
Q41	12/14 (85.7%)	5/5	2/2
Q43	3/14 (21.4%)	7/7	0/0
Q45	10/14 (71.4%)	4/4	3/3
Q47	5/14 (35.7%)	7/7	0/0
Q48	3/14 (21.4%)	3/3	0/4
Q49	9/14 (64.3%)	7/7	0/0
Q54	5/14 (35.7%)	7/7	0/0

Decomposition. For each after-window governed Claude answer whose question has before-congress data (v4 cycles and 27 Sep), naming (0/1) minus that question's before-congress naming rate; summed over answers that cite P61 and over answers that do not. Total excess: 80.8 namings over 378 after-window answers (+21.4pp). From answers citing the page: 83.8 (103.8% of the excess). From answers not citing it: -3.1 (2.4 in questions that drew the page elsewhere, -5.5 in questions that never drew it).

What this shows. Whether Claude's rise is concentrated where the page was cited. The questions that drew the page named Wegovy at 40.3% before the congress (9 of 31 at 50% or more), against 58.0% for the other questions; after it, 94.1% in answers citing the page and 43.1% in their non-citing answers, against 52.8% in questions that never drew it. **What it does not show.** A cause. This is a descriptive association, chosen after the data were seen: the page may be retrieved because a question already leads toward the medicine, and naming among citing answers is high partly because the page itself discusses it. Model or retrieval changes on or before 29 Sep would produce the same pattern.

11.5.12(d) Other engines, and Perplexity's 7 Oct step

Other engines citing the page: none.

Perplexity, UAE, governed answers: registered (brand or molecule), brand-only and molecule-only namings:

Cycles	N	Registered	Brand-only	Molecule only
6 Oct	54	17 (31.5%)	13 (24.1%)	4 (7.4%)
7 Oct	54	40 (74.1%)	31 (57.4%)	9 (16.7%)
4 and 6 Oct (pre-step)	108	40 (37.0%)	34 (31.5%)	6 (5.6%)
7–11 Oct (post-step)	270	203 (75.2%)	162 (60.0%)	41 (15.2%)

Step 6 → 7 Oct: registered +42.6pp, brand-only +33.3pp, molecule only +9.3pp. Pre-step → post-step: registered +38.1pp, brand-only +28.5pp, molecule only +9.6pp.

Paired by question, 6 → 7 Oct: 54 pairs; 24 answers turned to naming (17 naming the brand, 7 the molecule only) and 1 turned away.

Source groups cited on 7 Oct: answers that turned to naming citing the group, and naming among all 7 Oct answers citing / not citing it:

Group	Turned-to-naming answers citing it	Share of all 7 Oct answers citing it	7 Oct citing: named / answers	7 Oct not citing: named / answers
other: market research	21/24 (87.5%)	85.2%	32/46	8/8
clinic / hospital / care provider	16/24 (66.7%)	53.7%	27/29	13/25
other: social or platform	15/24 (62.5%)	53.7%	22/29	18/25
consumer / retail (incl. pharmacy)	11/24 (45.8%)	55.6%	20/30	20/24
regulator / government health body	10/24 (41.7%)	31.5%	15/17	25/37
news / medical media	7/24 (29.2%)	38.9%	16/21	24/33
unclassified	5/24 (20.8%)	20.4%	8/11	32/43
journal / literature / trial registry	3/24 (12.5%)	22.2%	9/12	31/42
other: directory or platform	3/24 (12.5%)	14.8%	6/8	34/46
other: insurer	2/24 (8.3%)	3.7%	2/2	38/52
other: professional or patient organisation	2/24 (8.3%)	5.6%	3/3	37/51
other: government non health	0/24 (0.0%)	1.8%	1/1	39/53
other: reference	0/24 (0.0%)	7.4%	3/4	37/50

Top pages new on 7 Oct (not cited by Perplexity on 6 Oct), by answers that turned to naming: N01 [market research, non-regional domain, H017] — 10 turned / 21 citing; N02 [market research, non-regional domain, H033] — 10 turned / 17 citing; N03 [social or platform, non-regional domain, H034] — 9 turned / 16 citing; N04 [social or platform, non-regional domain, H034] — 9 turned / 13 citing; N05 [social or platform, non-regional domain, H034] — 7 turned / 12 citing; N06 [market research, non-regional domain, H019] — 5 turned / 11 citing; N07 [market research, non-regional domain, H017] — 4 turned / 8 citing; N08 [social or platform, non-regional domain, H034] — 4 turned / 5 citing; N09 [market research, non-regional domain, H018] — 3 turned / 14 citing; N10 [market research, non-regional domain, H021] — 3 turned / 5 citing.

Market-research sources and naming: pre-step, answers citing one 11/30 (36.7%), not citing 29/78 (37.2%); post-step, citing 174/240 (72.5%), not citing 29/30 (96.7%).

What this shows. No single page accounts for the step: the new page most present among the 24 answers that turned to naming is N01 ([market research, non-regional domain, H017]), in 10. The group most cited by those answers is other: market research (21 of 24, 87.5%), but 85.2% of all 7 Oct answers cite it, so it is about as frequent among the turned answers as overall; post-step, naming is 72.5% in answers citing market research and 96.7% in those that do not. The group most over-represented among the turned answers is clinic / hospital / care provider (66.7% of turned answers against 53.7% of all 7 Oct answers). Of the post-step rise, brand-only naming accounts for +28.5pp and molecule-only naming for +9.6pp. **What it does not show.** A cause: on 7 Oct Perplexity's whole source mix changed at once (item 1(c)), so no group can be separated from the others, and the comparison is post hoc. Nor can it separate the step from the change in the structure of Perplexity's responses observed on the same day in question sets outside the study (PN1, paper §6.2).

11.5.12(e) The E4 link

The joint-class efficacy figures (the D4 reason P3: joint or class-level figure without valid brand attribution) that the E4 hand check (docs/evidence/easd-2026-manual-codes.json) excluded are this page's efficacy-figure sentence S2, repeated. 35 UAE answers echo S2 (Claude 35; 35 of them cite the page); 29 carry the coded figure. Scripted category: label 29. Hand check: excluded 29 (reasons P3 29). Of all 30 hand-checked exclusions of this figure as a joint-class efficacy figure (P3), 29 sit in answers echoing S2. So one local page is the source of a recurring joint-class E4 exclusion. The E4 quarantine is unchanged.

11.5.13 Disclosure scan

Before any file was written, every output (this file, the JSON, the CSV and the use-context codes file) was scanned for: every cited host that is not a public entity (clinics, hospitals, pharmacies, retailers, local businesses, insurers, directories, academic hosts and unclassified hosts); the Note 04 clinic page and its host; any run of 40 or more characters of the two item 12 page sentences, and the page's fingerprint strings (close paraphrases of those sentences and its exact date, held as hashes); and every question of the three batteries, whole when shorter than 30 characters and otherwise in overlapping 30-character fragments. None was found. Public entities (regulators, journals, the congress, manufacturers, news and medical media, reference publishers, professional and patient organisations) are named; everything else is typed.

12. Known failures and limitations

1. E4 attribution failed audit and is quarantined.

- The full-label audit (D17) found 12/150 wrong attributions (8.0%; Wilson 95% 4.6–13.5%) after adjudication, exceeding the 5% threshold, and 11.3% (7.2–17.4%) on first read. The pilot's 4–8% range had already indicated a possible threshold failure.
- The full D4 read of the no-source and congress-window figures excluded 104 of 195 figures through 10 Oct (113 through 11 Oct).
- No efficacy-source comparison is published.

2. The cache incident.

- The 5 Oct UAE cycle took 45 of 60 questions per engine from cache and was excluded (D14).
- The UAE after window has seven admitted cycles, not eight.

3. EDE outage and carry-forward.

- On 8 Oct and again on 11 Oct, the Emirates Drug Establishment directory failed its semaglutide control (D15, D20).
- UAE status is established only to 2 Oct; every after-window UAE E3 code is provisional.
- Saudi status rests on the 5 Oct record. For the E3 agents it is "not established", because the SFDA register could not be read.

4. Unreadable pages.

- 947 of 5,268 new (series, URL) pairs could not be read.
- The source route read 68.5%–90.1% of new-URL citations per engine in the after window.
- Unreadable pages are unverified, never negative. They concentrate in Saudi pharmacy, retail and delivery pages, in one Singapore source (mims.com) and in some UAE news sites.
- A congress source on an unreadable page would be missed by the source route. The text route flags only answers carrying a congress or readout term, so an answer that states such a result without naming it would be missed too (§7.1.5).

5. Historical page uncertainty.

- Pages were fetched on 10–11 Oct, not when cited.
- A later fetch does not establish what a page said at citation. Shared wording with an answer corroborates textual overlap only (§7.1.7).

- Pages cited before the congress and later updated are not re-read (detection limit).
6. **Model-assisted coding by one model family.**
 - First coding, adjudication and the blind second pass were all performed by Anthropic Claude models.
 - The second pass measures same-family consistency (E3 21/21, E5 κ 0.9768, E3-i κ 0.8663), not independent clinical validation.
 - Human reading covers a blind sample of 50 interpretive codes, read by the author as a single reader (§9.6). Naming was not human-coded.
 - The engines under study include a Claude model (claude-haiku-4-5). The coder is a different Claude model, but the shared family is a potential source of correlated reading.
 7. **The diaTribe miss and correction.**
 - The first ledger entry recorded TRIUMPH-2 only in its note, not in its readouts, because the term screen cannot see an unnamed trial.
 - An external check found it.
 - The correction keeps the earlier label in *revisions*, reports TRIUMPH-2 as unverified at citation, and leaves every registered cell unchanged (D18, D19).
 - Other unnamed readout content elsewhere cannot be excluded by the screen.
 8. **The date move (D20).** The analysis ran on 11 Oct rather than 12 Oct. The window could admit no further cycle. If any cycle had appeared before 23:59, it would be reported as outside the analysis.
 9. **Registration timing.** The protocol was registered after the congress readouts were presented and after the UAE during-window cycle of 29 Sep had been read (§1.3). The bands and decision rule were fixed before any after-window cycle existed, but not before all during-window data.
 10. **Post-registration readings.** D16's readings I1-I5 were written after preliminary after-window results were visible, though before any rate was compared with a band. The bridged-band construction (I1) determines the Claude verdict, whose margin is +0.395pp.
 11. **Instrument change.** The v4→v5 offsets rest on one bridge per series. The bridge interval widens each band end, and the Claude verdict depends on the stored upper bound (0.1173).
 12. **Same-question correlation.** The registered decision treats the pooled rate as fixed. The clustered intervals show that three UAE verdicts are borderline.
 13. **Saudi arm.** It is post-registration throughout. No engine's naming of teplizumab (Tzield) lies outside its bridged band, and no answer names an agent in the E3 set. E3-i is exploratory, sampled (cap 60) and compared only with its own pre-congress comparator; the clustered intervals overlap heavily. Nothing in this arm is a congress effect.
 14. **OCTANE.** It is post-registration (D3). Its programme entry was verified on 11 Oct, after the freeze, and its latency origin corrected (D21 item 3; §7.1.1).
 15. **Class (b) has no date condition (D21 item 8).** §4 defines class (b) as "a journal publication of a readout-set trial" with no date limit. Some readout-set trials were published before the congress (REIMAGINE 1 in June 2026; ACHIEVE-2, -3 and -5), so a new citation of any such paper in the congress or after window would qualify although it carries nothing first disclosed at EASD. D21 records the check made on 11 Oct: no readable new URL in any window is a readout-set publication, and the unreadable journal pages either name an unrelated article in their URL or, for ScienceDirect items without one, resolve through Crossref to unrelated articles. The gap has no effect on these results. It is a design limitation: a positive class (b) count would not by itself distinguish readouts first disclosed at EASD from readouts EASD re-presented.
 16. **The bridge self-check (D23, corrected by D24).** Raw three-run same-day ranges were narrower than the multi-cycle ranges. After a post hoc adjustment for the different numbers of runs, four of eight UAE/Saudi engine-market combinations exceeded the chosen between-day-variation threshold. The original range comparison alone does not separate within-day and between-day variation. The check was registered but reported only after the final run.
 17. **Errata after the freeze (D21).** The supplement audit found defects in the frozen record: a mixed Saudi E3 denominator, a quoted question, an unverified OCTANE origin, mis-stated times and counts, and stale notes. Each was corrected at source and every output regenerated on the same pull. D21 records that no registered verdict or registered result changed. The corrections were made after the analysis had been seen.
 18. **Registered §5 analyses reported late (D24).** The three descriptive source analyses registered in §5 were not run by the frozen analysis, and no deviation recorded the omission until D24. They are reported late in §11.5.1.
 19. **A single local page is associated with Claude's UAE rise (D24, §11.5.12; descriptive, post hoc).** This is an association, not a causal estimate.
 - A local commercial blog post (UAE), August 2026, before the congress, first appeared in Claude's resolved study citations on 29 Sep (among the resolved citations examined; pre-congress Gemini destinations were unresolved). It was then cited in 19–24 of 60 Claude UAE answers in every after-window cycle.
 - Claude named the medicine in 143/152 after-window governed answers citing it (94.1%), against 113/226 (50.0%) not citing it.
 - The citing answers echo two of the page's sentences; no non-citing answer after 28 Sep does.
 - The page may be retrieved because a question already leads toward the medicine.
 20. **Band-rule reading sensitivity (D24, §11.5.2).** One verdict flips under each of two alternative readings of the registered band rule (Perplexity UAE under the half-width reading; Claude UAE under the full-width reading). The deciding reading, I1, was fixed after counts were displayed.
 21. **Per-cycle and pooled readings differ (D24, §11.5.4).**
 - Perplexity's after-window naming lies outside its bridged band in 6 of 7 cycles, in both directions; its pooled "inside" verdict averages a low segment and a high one.
 - Claude's lies outside in 3 of 7 cycles.
 - Cycle-level change net of the offset (§11.5.3): Claude UAE +16.8pp (95% CI +7.2 to +26.4) and Perplexity UAE +12.2pp (−3.0 to +27.4). This is a different estimand, assuming independent cycles.
 22. **Text-route recall (D24, §11.5.11(b)).** Two answers stating the OCTANE result without naming it were missed by the text route; one is counted by no E1 route, which makes the registered-market either-route tally seven of 2,408 rather than six (§10.2). Recall for unnamed readout content is unknown.
 23. **The attribution test had little power (D24, §11.5.8).** Under the frozen E1 classification, for three of the four engine-arms with qualifying answers, no split could have reached $p < 0.05$. Under the corrected classification (§10.2), UAE Gemini's minimum attainable p is 0.206, still above 0.05; no verdict changes.
 24. **Endpoints, not products.** The four engines are API endpoints with model-decided search (Perplexity always retrieves), and no user location is sent. Results describe these endpoints, not the consumer apps or Google AI Overviews.
 25. **Human validation is limited (§9.6).** Blinded human validation of the interpretive coding (50 sampled E3, E5 and E3-i judgements), by the author as a single reader, is a set of selected quality checks (17/17 ordinary-stratum and 30/33 risk-enriched codes agreed), not an estimate of overall coding accuracy. It is not independent or external validation. The naming measure was not human-coded; its re-scoring is a deterministic consistency check (§9.1).
 26. **Gemini (the publication's Gemini constraints, §13.3).** Gemini results are reported in aggregate; verbatim outputs and grounding links are not redistributed. The four Gemini answer spans in the E3 ledger (§10.8) and one Gemini fragment in the second-pass table (§9.2) are withheld; their codes are kept.
 27. **Withheld sources and replication.**

Local-business identifiers are withheld from the publication and public evidence files under the study's disclosure policy. This restriction does not prevent independent researchers or AI systems from identifying or citing the same publicly accessible sources when conducting comparable measurements.

Independently discovering such a source is not exact replication: the question battery is withheld (\$4.2).

28. Scope of inference. The study measures citations and answer text in four engines, three markets and fixed batteries. It does not measure prescribing, demand, sales or traffic.

13. Reproducibility

13.1 Freeze record

Source: docs/evidence/computed/easd-2026-freeze.json (repository commit 6d1b45d360c8d74cdd026cf7d2d7eb256e2d3e07, frozenAt 2026-10-12T09:50:57.567Z, problems: none).

Role	Path	Bytes	sha256	In repository	Committed as is
pull	<OS temp>/easd-2026-study/d11/pull-final.json	110898974	de98a60469b0e7b100110f16170b16760cc32fe333696eb4292bf48219eed265	no	n/a (local)
regulator-status	docs/evidence/computed/easd-2026-regulatory-status-2026-10-11.json	8961	0741fc6d9ecbaae3156718c3fd2b2e30253500626cb2eb33d04a305ec2c73ef7	yes	yes
regulator-status-saudi	docs/evidence/computed/easd-2026-regulatory-status-saudi-t1d-2026-10-05.json	34538	618e1cf86e419f4f4041f2e27de5ed811a4bfa7ef3c012602f351c9f27277953	yes	yes
e4-reference-values	docs/evidence/easd-2026-e4-reference-values.json	18544	e7fe74de81b15f7969e36a7dc0308351ace710d07496195837eed601fde05a3d	yes	yes
manual-codes	docs/evidence/easd-2026-manual-codes.json	179357	b7d0cad0f0abf02210a29a6d0b445a99f0f73ac625f7565a726bc688e24558cf	yes	yes
saudi-manual-codes	docs/evidence/easd-2026-saudi-manual-codes.json	114138	65f0cfe666ccb6444567835ba802a39ec7e9600a367fde860f7198c9977bf531	yes	yes
final-coding	docs/evidence/computed/easd-2026-final-coding.json	1142618	4f773dbcf8b337864c13175107727ec3b976dc433238eacb4829292518508bd3	yes	yes
final-coding-ledger	docs/evidence/easd-2026-final-coding-ledger.md	378382	21c39cd219b65e9d3cde94e0f1829aa3302203094fe1ce31aae0ddc51be8d6f	yes	yes
saudi-final-coding	docs/evidence/computed/easd-2026-saudi-final-coding.json	230991	8cbaed40819bdc98cae8059124b2402f37af38684081e1cd4bbf438273c5b1d8	yes	yes
saudi-final-coding-ledger	docs/evidence/easd-2026-saudi-final-coding-ledger.md	115854	fc53ccd810a1d1f3c426125f89e82dd3e7287b9313e1ee6b4b076027c4569b6f	yes	yes
verify	docs/evidence/computed/easd-2026-verify.txt	2077	ccd68ef37cd91a4f2026edc3b06b95443da408cf760564bb39943f46b63389e1	yes	yes
bridge	docs/evidence/computed/easd-2026-bridge.json	13858	8286987d7d0581520acdcd8fc91252fe1cf2261c068a662be7d394a5e89ed146	yes	yes
e1-ledger	docs/evidence/easd-2026-e1-ledger.json	79056	07d8f49d5bfab27561d584575a82b1c7a4d8114f6cce1e9b4f0df93ad9f3ac8c	yes	yes
e1-manifest	docs/evidence/computed/easd-2026-e1-manifest.json	4842821	e6403e881f771f48d231aa8035efb9b33bb9dd2406deee7b6682580d699fad5e	yes	yes
e1	docs/evidence/computed/easd-2026-e1.json	107758	9384d569c49ca5908c8ef442663786d8db74518d35e4af44884d87a049f1de05	yes	yes
e1-md	docs/evidence/easd-2026-e1.md	22083	a4470a0a97763e2c2c2f120e0b8a4cb5cbe3a7367c85458ea4fbda826fe480ba	yes	yes
analysis	docs/evidence/computed/easd-2026-analysis.json	90163	18790cf6af0808ddd5f20d13a9f9aef9ebb193be359eb53e703346bb0ead39432	yes	yes
analysis-md	docs/evidence/easd-2026-analysis.md	13187	c1e7be813eccc5c0416eb87a1e0f827fa04406510be230a8e9850eb36befd034	yes	yes
second-pass	docs/evidence/easd-2026-second-pass.json	114510	4d547c4ec040743ecdda4ade595e83f48d99dc2478935a4ba86430f333dd5320	yes	yes
second-pass-builders	docs/evidence/easd-2026-second-pass-builders.py	7553	903046566334e935c74d4784f341584614122a5bedf541d0103be7be3fd954f8	yes	yes
e4-label-audit	docs/evidence/easd-2026-e4-label-audit.json	65077	359e2e96154152b72b9833cd1377eb7c45a22e59958d9c9c9e31036e5d0ebb78	yes	yes
t1d-agent-set	docs/evidence/easd-2026-t1d-agent-set.md	14321	37c16a012364a069b02144880460b515c3cb9f942475c83a4d977a8cdb27d4d8	yes	yes
baseline-coding	docs/evidence/computed/easd-2026-baseline-coding.json	759025	6ef1b64345721b75bc33ddf2faed40c96b709895472c280daa1457613167fe1b	yes	yes
saudi-baseline-coding	docs/evidence/computed/easd-2026-saudi-baseline-coding.json	93921	81dcc84ef768fd654af5bf98e0699ccfd4c7ba1a57a8c490574a3f7d92f7cc6	yes	yes
protocol	docs/design/easd-2026-evidence-to-answer-prereg.md	50392	204153f68887dc06102fa49f8d0bfb201c7cb0de1d6c1b27f19f3c366c087756	yes	yes
robustness	docs/evidence/computed/easd-2026-robustness.json	451636	5c6806bb5ca0897970381b5946e63c5ba2401776f9ca6a1dea2da99eb2cac124	yes	yes
robustness-md	docs/evidence/easd-2026-robustness.md	88893	8fe38bd1592ded04c9ae627af69ece59a4672053b8fbd1e7fa5fd15d3d0578fe	yes	yes
per-answer-outcomes	docs/evidence/computed/easd-2026-per-answer-outcomes.csv	1338403	68049e4c91a3c7d5f161609ae3c935325bc8ffaa5b226fa7eabab1d25f0fc6c7	yes	yes

Role	Path	Bytes	sha256	In repository	Committed as is
saudi-e3i-use-context	docs/evidence/easd-2026-saudi-e3i-use-context.json	61761	069d1d504309fd8a8e44bad4520636415066f32a91d367d2daf00bce1275109c	yes	yes
gemini-redirect-cache	<OS temp>/easd-2026-study/e1/gemini-resolved.json	9714719	0aa4612beaac5d2f2113ff2bf41129c3b7fb60e89159a65efdb11fbec3487427	no	n/a (local)
human-validation-protocol	docs/research/easd-2026/human-validation-protocol.md	19657	d5c3053026ddef913ed793aa25391fb6df52bb204f005244eb5c3041a42b6fa5	yes	yes
human-validation-results	docs/research/easd-2026/human-validation-results.md	9511	3c5b59005100d334f2f2f7fba3df743eef77af6d8fcc6eb a0c48f2b3b11a5d8e	yes	yes
cycle-codes-internal	docs/evidence/computed/easd-2026-cycle-codes.json	8848	a8f53c8b96ab06b78df4d74340db939bb49ca55d9f0f819ad5a2ad0cfbdfc619	yes	yes

Pull metadata (files[0].pull): pulledAt 2026-10-11T16:56:27.066Z, from 2026-08-22 to 2026-10-11, 67 cycles (uae-glp1 23, singapore-sglt 24, saudi-t1d 20).

Scripts:

Script	sha256	Committed as is
scripts/easd-2026-study/analyze.ts	18dd8cfab3ee6d08f46c4d769be8d1cdf2d8f714faf7510f47c8a4c4ff3dc602	yes
scripts/easd-2026-study/bridge.ts	8ac5410e53b7c8f78da6f4f596274d300d490062fc501efdd618977dab99ddb	yes
scripts/easd-2026-study/code.ts	d149288c9200d9823db3ea0db32a09225feb2741b56f2e84fed9d69ff4b06aa7	yes
scripts/easd-2026-study/cycles.ts	ce13be2df736649166dabb46d3a5d1db8c682211f730a344c8a0eef3e8301a4d	yes
scripts/easd-2026-study/e1.ts	fb97834c129e836fcfd4a60d31b3968f0e8530552f68f337c0246557bbc0c9	yes
scripts/easd-2026-study/evidence-package.ts	b348d71c1eb622e0acc15472813a965be406f986a271b2878ab129bff4c6df	yes
scripts/easd-2026-study/figures.ts	a0829f3d4cc3de879c1436aa598dbba40bd351ee1a4f20939aefda3c1058c4bf	yes
scripts/easd-2026-study/freeze.ts	529fd4bb3d2d59da833e8c1e21afb1d3d7c64ebf37413bf304058e65668b0760	yes
scripts/easd-2026-study/lib.ts	8d3bc2e40a9ea1c5aeac0031bcd7baeffefd74bf43d3b1aaa247d84fb2ce247	yes
scripts/easd-2026-study/pull.ts	b4fad3b11be312f57236467cade17fc92d645ef10a3a612b8106a45f3bb46492	yes
scripts/easd-2026-study/regulatory-status-t1d.ts	a6fb6436df48044fae1764b9eba60561ef3f5f6d7a41cb94db5bb5865de605d5	yes
scripts/easd-2026-study/regulatory-status.ts	a193e0bf03108f3230f2e9b5f8f218d6cc10b2f6bae52fb1cd605202ee76ae5f	yes
scripts/easd-2026-study/release-scan.ts	8aeba6af218ea34bdaff0273de84f9c7400d239cee9d569ec8ee7c027f867153	yes
scripts/easd-2026-study/robustness.ts	6819f10bb4d31a3b1e4617aff122e70b978be9ba5302941ee6aa5e41b8644af9	yes
scripts/easd-2026-study/verify.ts	a4e6d85676e1b1de52b675901d571afbc6973ef7da6557f43d4a34098080ba9a	yes

13.2 Commands

From the scripts' usage headers, with the file names of the final run as recorded in each output's inputs. All scripts run from the repository root at the commit the freeze record names.

Which of these scripts are published. The freeze record hashes 15 scripts, including the evidence-package builder. Two approved, sanitised verification scripts are released in the package's code/ (lib.ts and verify.ts); the other 13 are withheld, including evidence-package.ts (implementation proprietary; the computed outputs and specifications are published): analyze.ts, bridge.ts, code.ts, cycles.ts, e1.ts, evidence-package.ts, figures.ts, freeze.ts, pull.ts, regulatory-status-t1d.ts, regulatory-status.ts, release-scan.ts and robustness.ts. The commands below document how the frozen outputs were produced. Only the steps that use the released scripts can be re-run from the evidence package.

- pull.ts, cycles.ts, bridge.ts and analyze.ts read the production database, read-only: credentials in .env.local.
- regulatory-status*.ts and e1.ts --stage resolve|screen read public web pages.
- code.ts, e1.ts --stage tally and freeze.ts read local files only, the held pull among them. verify.ts reads only the committed coding outputs.

```
# 1. Pull (read-only; writes question and answer text to a local file outside the repository)
npx tsx scripts/easd-2026-study/pull.ts --from 2026-08-22 --to 2026-10-11 --out <pull.json>
#   Match the result against the freeze record: sha256 de98a60469b0e7b100110f16170b16760cc32fe333696eb4292bf48219eed265

# 2. Regulator snapshots (public registers; never scheduled)
npx tsx scripts/easd-2026-study/regulatory-status.ts --out <new.json> \
  --previous docs/evidence/computed/easd-2026-regulatory-status-2026-10-02.json
npx tsx scripts/easd-2026-study/regulatory-status-t1d.ts \
  --out docs/evidence/computed/easd-2026-regulatory-status-saudi-t1d-2026-10-05.json

# 3. Coding (E3, E4, E5; Saudi E3, E3-i); exit code 2 if anything is uncoded
npx tsx scripts/easd-2026-study/code.ts --phase full --pull <pull.json> \
  --status docs/evidence/computed/easd-2026-regulatory-status-2026-10-11.json \
  --refs docs/evidence/easd-2026-e4-reference-values.json \
  --manual docs/evidence/easd-2026-manual-codes.json \
  --out-json docs/evidence/computed/easd-2026-final-coding.json \
  --out-md docs/evidence/easd-2026-final-coding-ledger.md \
  --saudi-status docs/evidence/computed/easd-2026-regulatory-status-saudi-t1d-2026-10-05.json \
  --saudi-manual docs/evidence/easd-2026-saudi-manual-codes.json \
  --saudi-out-json docs/evidence/computed/easd-2026-saudi-final-coding.json \
  --saudi-out-md docs/evidence/easd-2026-saudi-final-coding-ledger.md
npx tsx scripts/easd-2026-study/verify.ts --json docs/evidence/computed/easd-2026-final-coding.json --md docs/evidence/easd-2026-final-coding-ledger.md --
saudi-json docs/evidence/computed/easd-2026-saudi-final-coding.json --saudi-md docs/evidence/easd-2026-saudi-final-coding-ledger.md >
docs/evidence/computed/easd-2026-verify.txt

# 4. Bridge (paid; already run; never re-run to obtain a different offset, 56)
#   npx tsx --conditions=react-server scripts/easd-2026-study/bridge.ts --series uae-glp1 --validate-only
#   (--validate-only re-scores the source cycle and makes no paid call)

# 5. E1
```

```

npx tsx scripts/easd-2026-study/e1.ts --stage resolve --pull <pull.json>
npx tsx scripts/easd-2026-study/e1.ts --stage screen --pull <pull.json> --manifest docs/evidence/computed/easd-2026-e1-manifest.json
npx tsx scripts/easd-2026-study/e1.ts --stage tally --pull <pull.json> \
  --ledger docs/evidence/computed/easd-2026-e1-ledger.json --manifest docs/evidence/computed/easd-2026-e1-manifest.json \
  --admissions docs/evidence/computed/easd-2026-final-coding.json,docs/evidence/computed/easd-2026-saudi-final-coding.json \
  --out-json docs/evidence/computed/easd-2026-e1.json --out-md docs/evidence/easd-2026-e1.md

# 6. Analysis (re-scores every answer through the production scorer and checks the governed counts)
npx tsx --conditions=react-server scripts/easd-2026-study/analyze.ts --pull <pull.json> \
  --coding docs/evidence/computed/easd-2026-final-coding.json \
  --saudi docs/evidence/computed/easd-2026-saudi-final-coding.json \
  --bridge docs/evidence/computed/easd-2026-bridge.json --e1 docs/evidence/computed/easd-2026-e1.json \
  --out-json docs/evidence/computed/easd-2026-analysis.json --out-md docs/evidence/easd-2026-analysis.md

# 7. Post-registration robustness (D24; reproduces every registered count and verdict first)
npx tsx --conditions=react-server scripts/easd-2026-study/robustness.ts \
  --pull <pull-final.json> --coding docs/evidence/computed/easd-2026-final-coding.json \
  --saudi docs/evidence/computed/easd-2026-saudi-final-coding.json \
  --bridge docs/evidence/computed/easd-2026-bridge.json --e1 docs/evidence/computed/easd-2026-e1.json \
  --manifest docs/evidence/computed/easd-2026-e1-manifest.json \
  --out-json docs/evidence/computed/easd-2026-robustness.json --out-md docs/evidence/easd-2026-robustness.md \
  --out-per-answer docs/evidence/computed/easd-2026-per-answer-outcomes.csv

# 8. Freeze
npx tsx scripts/easd-2026-study/freeze.ts --out docs/evidence/computed/easd-2026-freeze.json \
  --file pull=<pull.json> --file analysis=docs/evidence/computed/easd-2026-analysis.json [--file role=path ...]

```

Notes on reproduction:

- **Determinism.** `code.ts`, `e1.ts` `--stage tally` and `analyze.ts` write no wall-clock field (`deterministic: true`). Bootstraps are seeded (§5.2, §8.4). Given the same pull, the same committed inputs and access to the same stored scan inputs, the outputs are byte-reproducible.
- **Stage 3 and `--phase`.** The final coding outputs record phase: "full". `--phase` defaults to `baseline`, which caps each series at its pre-congress cutoff. The baseline outputs (`easd-2026-baseline-coding.json`, `easd-2026-saudi-baseline-coding.json`) were produced with `--phase baseline` and the 2 Oct snapshot.
- **Stage 5 depends on mutable inputs.** `e1.ts` `--stage resolve|screen` reads live web pages and caches them in the OS temp directory. A rerun on a later date will not fetch identical pages. The committed manifest and ledger are the frozen record of what was read.
- **Usage headers that name other files.** The usage header in `analyze.ts` names `easd-2026-after-coding.json` and `easd-2026-saudi-after-coding.json`. The final run used the `final-coding` files, as its `inputs` field records.

13.3 Data availability: the public evidence package

Data availability follows the author's publication decisions, set out in this section.

- AlphaCitation publishes a public evidence package with the paper and this supplement (package version v1.0.2; v1.0 and v1.0.1 are archived, paper version history):
 - the paper: <https://alphacitation.com/blog/does-new-medical-evidence-change-ai-treatment-answers>
 - this supplement: <https://alphacitation.com/blog/does-new-medical-evidence-change-ai-treatment-answers/methodology>
 - the evidence package: <https://alphacitation.com/blog/does-new-medical-evidence-change-ai-treatment-answers/evidence>
- Its `README.md` maps every folder to the paper and to this supplement. `DATA-DICTIONARY.md`, `MANIFEST.json` and `REDACTION-LOG.md` give the data dictionary, the file manifest with hashes, and every redaction rule with its count.
- Before release, the package passed a dedicated disclosure audit covering:
 1. leakage of the exact question battery through scripts, identifiers, answer excerpts, filenames, metadata or combinations of fields;
 2. clinic, pharmacy and local-business identifiers;
 3. API credentials, internal system details and commercially sensitive methodology;
 4. third-party copyright, data licensing and permitted redistribution;
 5. consistency between the public files, the frozen analysis, the paper and this supplement.
- Its gates and the release scan are recorded in its `README.md` and `REDACTION-LOG.md`.

Contents, each passed by the disclosure audit:

- the registered protocol and the complete dated deviation history;
- endpoint specifications;
- E1 and E2 outputs;
- bridge and baseline analyses;
- coding ledgers and manual codes, suitably redacted (opaque prompt identifiers and answer excerpts; no question text; no verbatim Gemini text);
- the E1 ledger and manifest, with clinic, pharmacy and local-business identifiers redacted;
- the second-pass reliability analysis;
- the E4 audit;
- regulatory records;
- figure data;
- verification results (including `easd-2026-verify.txt`);
- the reproducibility record (the freeze record, §13.1);
- `code/`: the two released verification scripts (§13.2).

The package `README` states, for every material published finding, its verification path, and which analyses are reproducible from public files and which require restricted data.

Availability. Public files support recalculation of the specified aggregate and per-answer-outcome analyses and inspection of selected redacted excerpts. Full answer-level recoding, exact prompt replication, production-scorer verification and the E1 tally require restricted inputs and/or unreleased code. Full raw answers are withheld for every engine.

From the commands in §13.2:

- **Public files only.**

- The released verifier, run on the public coding files, reproduces the frozen verifier output. The single command, from the package root, tested on this version:

```
npx tsx code/verify.ts --json coding/final/easd-2026-final-coding.json --md coding/final/easd-2026-final-coding-ledger.md --saudi-json coding/final/easd-2026-saudi-final-coding.json --saudi-md coding/final/easd-2026-saudi-final-coding-ledger.md
```

Its output is identical to `verification/easd-2026-verify.txt`.

- Every table in §10 that is copied from a published output can be recalculated from public files.
- The per-answer outcome table (§11.5.10; no question or answer text) is enough to recompute E2, readings I1 and I2, I3 under the frozen classification (the corrected I3 also needs the package's correction record, `robustness/easd-2026-e1-corrections.csv`), the clustered and per-cycle analyses, D24 items 2–8 and item 12's central split. Its columns are (searched marks an answer with at least one citation): `series`, `cycleSlug`, `cycleDate`, `window`, `engine`, `promptId`, `focused`, `inGovernedPopulation`, `namedRegistered`, `namedBrandOnly`, `searched`, `citationCount`, `e1Qualifying`, `citesUnverifiedNewPage`, `citesLocalPageP61` and `echoesLocalPageSentence`. The last two flag answers that cite the local page and answers that repeat S1 or S2 under the 40-character rule. Neither the column names nor the values identify the page.
- **Restricted data needed.** Stages that read the pull or the production database cannot be rerun from public files: the coding, the E1 tally, the analysis (which re-scores answers through the production scorer), and the bridge.

Source-identification policy (the amended publication policy).

- **Named:**

- major non-local organisations: established organisations operating nationally or internationally under their own name, outside the measured markets, such as journals and scientific publishers, medical and drug databases, academic institutions and eligible academic and non-profit hospitals, research organisations, news media, sponsors, international retailers, major pharmacy chains, and major commercial and healthcare companies. Publishers are named wherever they are based.
- public regulators and government bodies, wherever they are, including those in the measured markets.

- **Tokenised:**

- everything operated or headquartered in a measured market, except public bodies: local clinics, hospitals, pharmacies, laboratories, insurers, care providers and businesses;
- non-local entities not shown to be major organisations, including small clinics, pharmacies and practitioners.
- This follows the amended publication policy (§2.1), not the original registered boundary.
- Each withheld entity carries one typed token from the package's `redaction-token-map.json` (for example `[pharmacy-02]`). The same token stands for the same entity in every file of a version, and the full mapping is held in AlphaCitation's restricted archive.
- Tokens are stable within a package version, not across versions, so the tokens quoted here are those of v1.0.

- **Counts** (from the package's `REDACTION-LOG.md` and `redaction-token-map.json`): 2275 hostnames appear in the frozen files. 916 hostnames of 834 named entities keep their URLs (the Gemini-only pages apart). 1276 entities are withheld, each replaced by one typed token, and 49 platforms carry platform tokens. 572 non-local entities are withheld as not major organisations; 355 of those 572 “not a major organisation” entries are small businesses, practitioners, distributors, affiliates or personal sites (token map, reasons beginning “not a major organisation”). 28 pages of named entities have their title or path withheld (rule R1f; the pages are listed in the log).
- **Consistency.** One documented policy applies across all engines, windows and series. No source is identified or concealed according to the direction of a finding.
- **Cited, overlapping, causing.** A source being *cited* by an answer (a link in the answer's citations) is a different fact from its wording *overlapping* the answer's (shared wording between a captured answer and a later-fetched page, as in §7.1.7 and §11.5.12(b)). Overlap corroborates shared text only (§7.1.7). Both differ from a source *causing* an answer to change, which this study does not establish for any source. Associations between citing a source and naming (§11.5.12) are descriptive.
- **Further review.** Public source-level disclosures are reviewed for battery leakage and third-party redistribution limits.

The publication's Gemini constraints. Gemini results are reported in aggregate; verbatim outputs and grounding links are not redistributed.

- No grounding-redirect string appears in any public file.
- A page cited only by Gemini is not identified: its URL, publisher and title are withheld (in this supplement: four qualifying sources in §10.5).
- A page also cited by another engine keeps its URL under the source policy above.
- Search Suggestions were never captured, and none are published.

Replication of withheld sources.

Local-business identifiers are withheld from the publication and public evidence files under the study's disclosure policy. This restriction does not prevent independent researchers or AI systems from identifying or citing the same publicly accessible sources when conducting comparable measurements.

Independently discovering such a source is not exact replication: the question battery is withheld.

Licence and third-party material.

AlphaCitation is a digital product of Akunudo LLC-FZ. Akunudo LLC-FZ is the copyright holder and licensor of the original material identified in this publication. That material is licensed under the Creative Commons Attribution 4.0 International licence (CC BY 4.0), subject to the third-party exclusions stated below.

- **Licences**, granted by Akunudo LLC-FZ and set out asset by asset in the evidence package's `LICENCES.md`:
 - **CC BY 4.0:** this supplement, the paper, the executive brief, and the original published figures and exhibits.
 - **MIT:** the approved, sanitised research-verification scripts in the evidence package (`code/`).
 - **CC BY 4.0, where Akunudo LLC-FZ holds the redistribution rights:** the redacted evidence datasets. Any file or field derived from third-party material keeps its own rights and restrictions, and the publication's Gemini constraints apply.
 - **Third-party material** keeps its original rights and contractual restrictions.
 - **Proprietary and expressly excluded:** the active question battery, AlphaCitation's measurement instruments, production software, confidential methodology implementations and unreleased datasets.
 - Nothing beyond the approved public evidence package is released or licensed.

- **Third-party exclusions.** Except where separately identified, the original text, figures and compiled research outputs for which Akunudo LLC-FZ holds the relevant rights are licensed under CC BY 4.0. Quoted third-party text, third-party data fields, API/provider-restricted material and third-party marks are excluded from that grant unless their own licence expressly permits the stated reuse.
- Third-party extracts keep their sources' terms:
 - FDA data via openFDA (CC0).
 - Source: European Medicines Agency (medicines data; EPAR product information), ema.europa.eu, accessed October 2026.
 - Contains information from Listing of Registered Therapeutic Products accessed on 2 and 11 October 2026 from data.gov.sg, which is made available under the terms of the Singapore Open Data Licence version 1.0 (<https://data.gov.sg/open-data-licence>).
 - EDE, SFDA and EASD extracts are factual excerpts with their sources named, and imply no endorsement.
- AI-output excerpts are the study's API outputs, labelled by engine. No Gemini excerpt is reproduced (the publication's Gemini constraints).

Version identifiers inside frozen outputs. `easd-2026-robustness.json` and `easd-2026-robustness.md` record the freeze in force when they were computed (84a2214c, 2026-10-11T20:23:25Z). That record has since been superseded. Their frozen inputs are byte-identical to those in the current freeze record (§13.1), which also hashes the local Gemini redirect cache. The robustness JSON describes that cache as "not in the freeze record", which was true when it was written.

Correspondence. Scholarly correspondence, methodological critiques and correction requests: analyst@alphacitation.com (§9.7). This is separate from the commercial "Request an AlphaCitation briefing" route.

Never published:

- **The pull** (`pull-final.json`). It holds full question text and full raw answers. Full raw AI answers and the active question battery are never published (protocol §10; §4.2; §13.3). The pull is identified by its sha256 in §13.1, so any copy can be matched exactly.
- The unredacted E1 ledger and manifest. They are held by AlphaCitation and matched by the sha256 values in §13.1.
- **One frozen E4 answer span.** The frozen final-coding file (`easd-2026-final-coding.json`, E4 rows) contains one of the local page's sentences verbatim, as repeated in a Claude answer (§11.5.12(e)). It stays in the frozen file, whose hash the freeze record pins. The evidence package redacts it.
- The local Gemini redirect cache. It is hashed in the freeze record (D24) because the E1 tally and the robustness script read it.
- The local caches: resolved redirects, fetched pages, candidate lists, and the blind-pass worksheets, which carry question text.
- The production scorer's internals.

The question-instrument statement (also in §4.2):

Question-instrument availability. The exact question battery is withheld to preserve the integrity of AlphaCitation's longitudinal measurement instrument. Public disclosure of the active questions could enable content to be optimized specifically against the instrument, potentially influencing future observations and compromising comparability over time.

To support scientific scrutiny, the study reports its question taxonomy, sampling design, endpoint definitions, coding procedures, analytical methods and reproducibility controls. The complete instrument is preserved in the frozen study record.

This restriction represents a departure from CHART Item 5b and limits independent replication of the exact data-collection procedure.

Withholding the battery limits independent replication of the exact data collection. It does not limit the disclosure of methods, rules, outputs or verification paths.

The protocol states no data-access procedure, so this supplement states none. A reader who needs to examine held material may write to analyst@alphacitation.com. Requests for clarification and methodological critique are welcome under the open-scrutiny statement in §9.7.

13.4 Version history

Version	Date	Change
v1.0	October 2026	First release. Analysis of 11 Oct 2026 (d1fccaa9), with the D21 errata regenerated on the same pull (b2c52188)