Keiji AI

For Real-World Data and Evidence Teams

TrialMind for feasibility, comparative effectiveness, treatment patterns, and safety — natural-language cohort extraction that runs where your data lives, with a reproducible audit trail on every cohort definition, funnel, and analysis.

Keiji AI Inc · keiji.ai · contact@keiji.ai

01

What we hear from RWE leaders

Feasibility takes weeks to months of data engineering before any analysis can begin, so go/no-go decisions arrive late — often after the decision has effectively been made without them. The majority of analyst time goes to cleaning and joining data rather than generating insight. Claims, EHR, and registry sources do not join cleanly, and each new question restarts the engineering from something close to zero.

Meanwhile FDA, EMA, and PMDA keep widening their acceptance of real-world evidence in submissions. The same team, with the same headcount, is now being asked for output that must withstand regulatory scrutiny rather than internal scrutiny. The bar moved; the capacity did not.

And the remit keeps widening with it. The function that was asked for feasibility counts is now also asked for comparative effectiveness where no trial exists, safety commitments with dates attached, and the burden-of-illness picture medical affairs needs — often for the same molecule, from the same data, by the same four people.

02

Where TrialMind fits

Natural-language cohort extraction turns a feasibility question into an answer in minutes rather than weeks, without a data engineering cycle in between. The epidemiologist who has the question runs the analysis.

TrialMind reasons across the sources you already license rather than requiring another data purchase, and it **runs where your data lives** — raw records never move. That last point is usually what determines whether a project survives privacy and governance review.

The same platform carries the analysis past the cohort. A comparative question gets the emulated protocol, the adjustment, the balance diagnostics, and the sensitivity analysis in one traceable chain — so the answer arrives with the defense already attached rather than assembled afterward.

> **Every cohort definition, attrition funnel, and survival analysis carries a reproducible audit trail. That is what lets the output go into a submission rather than only into a slide.**

03

Seven workflows we will work on with you

WORKFLOW	WHEN IT COMES UP	WHAT YOU GET
Feasibility with an attrition funnel	Clinical asks "how many patients like this exist?" with an unrealistic deadline	An eligible-patient count plus the funnel showing how each criterion moved it — with the code
Comparative effectiveness by target trial emulation	A comparative question no trial will answer — including a single-arm study needing an external control	The emulated protocol stated explicitly, then executed: propensity or IPTW adjustment, balance diagnostics, and the sensitivity analysis a reviewer will ask for
Treatment patterns and drug utilization in routine care	Post-launch — who is getting this, in what sequence, and instead of what	Adoption, switching, sequencing, and adherence by disease, line, and biomarker status. In oncology this is line-of-therapy reconstruction, with every rule made explicit and applied the same way each time
Burden of illness and unmet need	Medical affairs needs the disease characterized before anything else can be argued	Incidence, prevalence, treatment patterns, outcomes gaps, and utilization — assembled as one picture rather than five commissioned studies
Safety signal evaluation and post-marketing commitments	A signal appears, or a PMR/PMC has a filed protocol and a fixed date	Longitudinal and disproportionality analyses, and committed studies executed to the agreed protocol with the audit trail the commitment requires
Eligibility stress-test and population generalizability	Protocol design before criteria are fixed, or a Diversity Action Plan	Per-criterion cohort impact, plus your planned population compared against the real-world disease population by stratum
Standing cohort refreshed on each data cut	Quarterly refresh of a definition already agreed	The same definition re-run, with a diff against the prior cut

Where we can start: feasibility with an attrition funnel. The speed is the headline, but the funnel is what makes it credible — a number without the funnel is just a faster guess.

The row that matters most is the second one. An external control arm is the highest-stakes instance of a target trial emulation, not a separate thing — so the same method that defends a single-arm submission also answers the comparative questions you will never run a trial for. That is one capability serving both halves of your remit rather than a pre-trial tool.

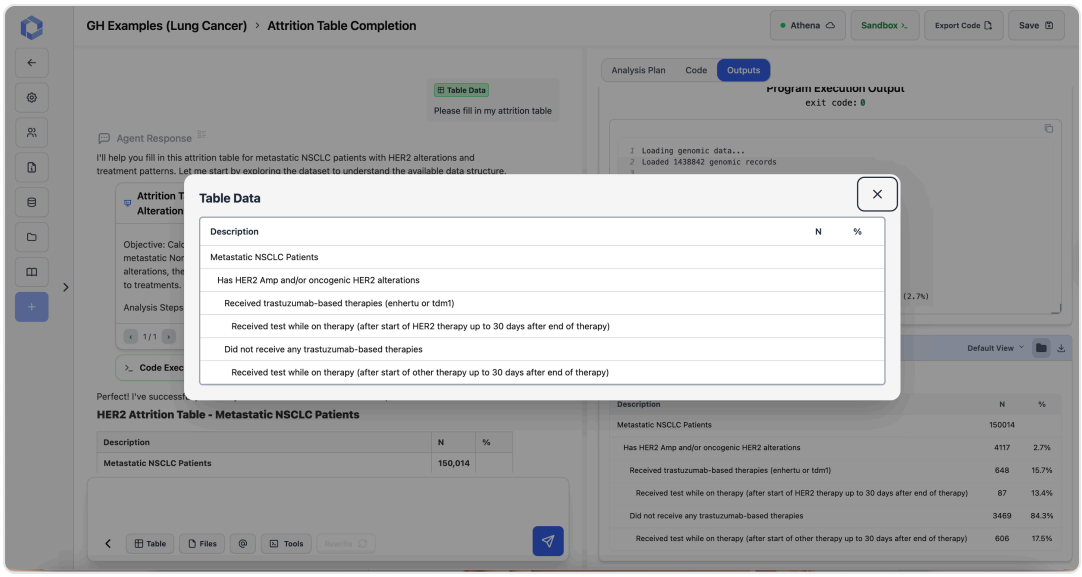
Before any of it, a data-fitness question: whether the sources you license can actually support the claim you need to make, which variables are missing or proxied, and what limitation survives. That is scoping work, not a deliverable, and we would rather do it in the first week than discover it in review.

04

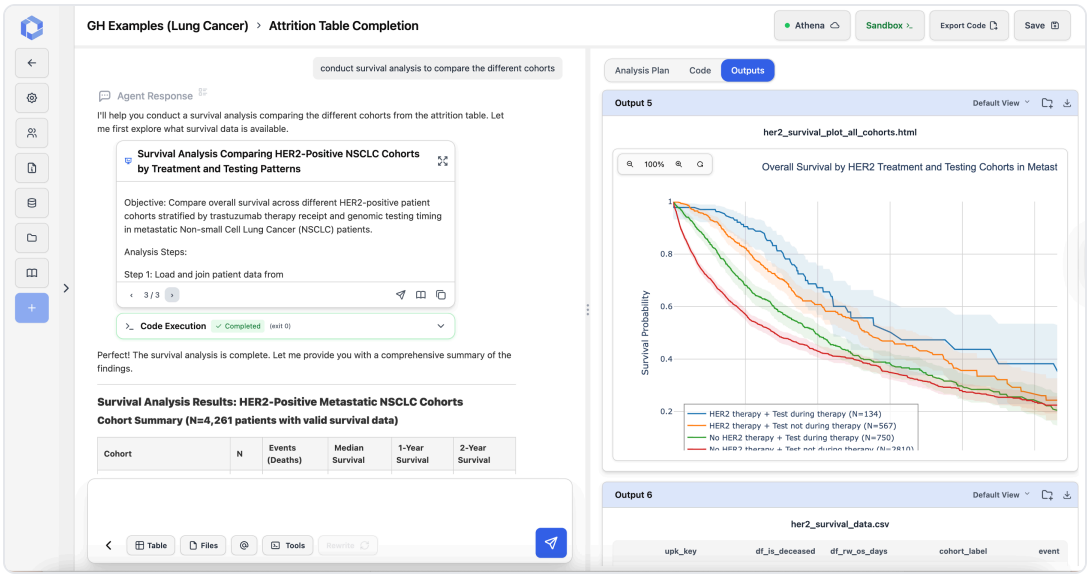
One workflow, end to end

HER2-altered metastatic NSCLC — from cohort definition to survival comparison, in one session

Step 1 — Define the real-world cohort and the analysis question. Specify the clinical population of interest (here, metastatic NSCLC with HER2 alterations) and the analytical objective — treatment patterns, testing behavior, or survival outcomes. The attrition table comes back with counts and percentages at every step, so the cohort can be inspected before anything is analyzed.



Step 2 — Run the survival analysis. Time-to-event analyses — here Kaplan–Meier overall survival — run across the cohorts with consistent definitions of index date, censoring, and follow-up; plan, code, and output are all retained.



05

What we would say about ourselves

> **Eight claims, and each one is a thing you can check in the first week rather than take on faith.**

01
WHERE IT RUNS

Data-neutral. We work across the sources you already license, so adopting us does not require another data purchase or an internal decision about which vendor wins.

Runs where the data lives. Raw records stay in your environment — which is what makes privacy and governance review passable.

Integrates with the data estate already in place — OMOP, claims, registry, and EHR sources alongside the SAS, R, Python, Spotfire, and Tableau stack your analysts work in.

02
WHO DOES THE WORK

Natural-language cohort extraction puts the analysis in the hands of the epidemiologist who has the question, with no data engineering cycle in between.

Bring your own code. The SQL, R, Python, and cohort definitions your team has already written and validated import as they are — they become part of how the agents work rather than something to be replaced.

Programmable and reproducible. Call TrialMind from your own scripts, including through Claude Code, and save a study so it re-runs on the next data cut and hands back a comparable result.

03
WHAT HOLDS UP UNDER REVIEW

Reproducible cohort definitions and analyses with a full audit trail, suitable to support regulatory use as FDA, EMA, and PMDA widen RWE acceptance.

Methods published in the peer-reviewed literature, so the approach can be defended by citation rather than by assertion.

06

Why this is defensible

Four things supporting the workflows, in the order they matter

Secure by design

The AI control plane is separate from the data plane. **Raw records never leave your environment**; the control plane operates on metadata — an architectural property rather than a policy commitment. We have passed enterprise security reviews at Guardant Health and Regeneron; the SOC 2 report, certificates, and completed vendor questionnaires are available on request.

SOC 2 Type II | ISO 27001 | HIPAA compliant · BAAs in place

Models built on clinical research, not adapted to it

Domain-specific models outperform general-purpose ones on clinical tasks, and the comparison is published rather than asserted. Test it against a cohort you have already built — that comparison is the actual pitch.

LEADS — literature mining trained on 21,335 systematic reviews and 453,625 citations; all-round improvement over GPT-4o on the paper's benchmark tasks (*Nature Communications*, 2025).

TrialGPT — 40% average time saving on patient–trial matching, deployed at the NIH (*Nature Communications*, 2024).

Trial2Vec — trial retrieval improved 7% absolute over the best baseline, 16% over the state-of-the-art medical model (*EMNLP Findings*, 2022).

A data pipeline behind the cohort

1.3M+
clinical trial publications

850K+
trial protocols across 14 registries

150K+
patient notes

The value is not the size of the index — it is that a cohort definition, an endpoint, or a comparator choice can be checked against how comparable studies did it, without leaving the analysis.

Regulated-grade access and audit

Role-based access is enforced at the reasoning layer, not bolted on at the interface — the agent cannot reason over data the user is not entitled to see, regardless of what any prompt asks for. Every output carries an immutable audit trail with a human checkpoint: actions, data changes, and system events logged, electronic signature capability built to 21 CFR Part 11 signature requirements, encryption at rest and in transit, and ALCOA+ throughout. Formal computer-system validation is a separate track, and we would rather tell you where that stands than let the controls imply it.

07

Three ways to work with us

The right model depends on where your team already is rather than on what we would prefer to sell.

1 License the platform — your team runs it

TrialMind deployed into your environment, operated by your epidemiologists and data scientists. **For functions that want to own and operate their own tooling** but need domain depth and traceability a general-purpose platform does not carry.

YOU GET	The platform, fitted to your data model — OMOP, claims, registry, or your own schema — and to the cohort conventions your team already uses
YOU OWN	The cohorts, the generated code, and the validated study workflows you build on it
COMMERCIALLY	Roughly 1–3 months to fit the agents to your data model; a one-time integration cost plus an ongoing licence

2 Delivered as a service — we produce the study

You do not buy software. You buy the output: a feasibility package, an external control arm with its balance diagnostics, a comparative effectiveness analysis. **For teams under a date they cannot move** — not a CRO relationship in a new wrapper, but a specific study done by the team that built the tooling.

YOU GET	The finished study — protocol, cohort definition, executed code, results, and the audit trail intact
YOU OWN	The deliverable and everything supporting it
COMMERCIALLY	Priced per deliverable, per study, or per milestone — an outcome, not a seat count

3 Build with us — your team owns it

For functions that have already decided to build. You have the engineers and the mandate; what is missing is clinical development grounding — what a defensible cohort definition requires, and where an agent's output stops holding up under audit.

YOU GET	Custom agents and validated pipelines on your stack, plus 100+ clinical MCP tools
YOU OWN	The pipelines, the skills, and the capability. Knowledge transfer is the deliverable, not a dependency
SCOPE	A defined scope with a delivery date — never open-ended staffing — at scoped project pricing

Most teams start narrower than they end. Moving between models is expected.

08

Relevant proof

Guardant Health	50+ data scientists and technical sales staff in production on oncology feasibility and real-world evidence. RWD feasibility studies moved from 6–8 weeks to 3–5 days, with roughly 40% less dependency on external analytics and about 3× analyst throughput.
Regeneron	Natural-language cohort extraction from real-world data, integrated via MCP alongside existing clinical systems.
Mass General Brigham and Beth Israel Deaconess	Clinical researchers running cohort analysis and OMOP-integrated studies themselves, in a secure sandbox with no special environment setup, on analyses that were previously infeasible without dedicated infrastructure.
Published result — DSWizard Nature Biomedical Engineering, 2026	A multi-step analysis workflow — dimension reduction, subtyping, differential expression, pathway enrichment, literature grounding — that takes roughly three months by hand completed in about a day, with a human setting the target and reviewing the results. That division of labour is the one we are proposing here: you specify the study and check the work; the agent does the iteration.

09

Relevant publications

The methods under this page, mapped to the papers your team can read before trusting anything downstream of them.

WHAT IT SUPPORTS	PAPER
Agent-run analysis you can check	*DSWizard* — reliable multi-step data-science agents for biomedical research (*Nature Biomedical Engineering*, 2026) · *Can Large Language Models Replace Data Scientists in Clinical Research?* (*Nature Biomedical Engineering*, 2024)
Longitudinal EHR modeling	*HALO* — synthesis of high-dimensional longitudinal records (*Nature Communications*, 2023) · *PromptEHR* (*EMNLP*, 2022) · *SynTEG* — temporal structured EHR simulation (*JAMIA*, 2021)
Prediction on real-world populations	*MediTab* — EHR pre-training for zero-shot prediction on trial populations (*IJCAI*, 2024) · Evidence-driven spatiotemporal hospitalization prediction (*Nature Communications*, 2023)
Comparators without patient-level access	Individual patient data reconstruction from published trials — accepted at *Machine Learning for Healthcare* (MLHC) 2026
Population representativeness	Generative balancing for equity in medical machine learning (*npj Digital Medicine*, 2025)
Patient–cohort matching	*TrialGPT* (*Nature Communications*, 2024) — deployed at the NIH with a reported 40% average time saving

The two to send first are DSWizard, because it is the clearest statement of the division of labour we are proposing, and the MLHC reconstruction paper, because it is the one a methodologist will want to argue with.

The full record is in the companion page, *The Research Behind the Work* — every workflow mapped to the papers underneath it, with peer-reviewed work and preprints listed separately, and the places where we have no publication to offer stated plainly. Ask for it, or tell us which two or three matter most and we will send those.

Keiji AI

THE FIRST STEP

Bring one question you have already answered

You know the right answer; we run it and you check our work — including the funnel and the code, not just the count.

START HERE

One feasibility question, answered the slow way

We run it and you check our work — the eligible-patient count, the attrition funnel showing how each criterion moved it, and the generated code.

THEN

The comparative question you could not answer

Where no trial exists and the analysis stalled on confounding. We show the emulated protocol and the diagnostics before anyone commits to a study — so you are judging a method rather than a demo.

CONTACT

contact@keiji.ai

Keiji AI Inc · keiji.ai
Seattle, WA, USA