Keiji AI

TrialMind for Real-World Data and Evidence

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

STARTING POINT

What we hear from RWE leaders

Feasibility arrives late

Weeks to months of data engineering before any analysis begins. Go/no-go decisions land after the decision has effectively been made.

The bar keeps rising

FDA, EMA, and PMDA keep widening RWE acceptance. Output now has to withstand regulatory scrutiny, not internal scrutiny.

The remit keeps widening

Comparative effectiveness where no trial exists, safety commitments with dates attached, burden of illness — same molecule, same four people.

The bar moved. The capacity did not.

THE PLATFORM

Where TrialMind fits

A feasibility question becomes an answer in minutes — and the same platform carries the analysis past the cohort.

Natural-language cohort extraction

The epidemiologist who has the question runs the analysis — no data engineering cycle in between.

Runs where your data lives

Reasons across the sources you already license. Raw records never move, which is usually what decides governance review.

One traceable chain

Emulated protocol, adjustment, balance diagnostics, sensitivity analysis — the answer arrives with its defense attached.

WHY IT MATTERS

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.

01

The workflows

Seven we would work on with you, where we would start, and one of them end to end.

SCOPE

Seven workflows we would work on with you

WORKFLOW	WHAT YOU GET
Feasibility with an attrition funnel	Eligible-patient count, plus the funnel showing how each criterion moved it
Comparative effectiveness by target trial emulation	Emulated protocol, then IPTW adjustment, balance diagnostics, sensitivity analysis
Treatment patterns and drug utilization	Adoption, switching, sequencing, adherence — by line and biomarker status
Burden of illness and unmet need	Incidence, prevalence, outcomes gaps, and utilization as one picture
Safety signals and post-marketing commitments	Longitudinal and disproportionality analyses, run to the filed protocol
Eligibility stress-test and generalizability	Per-criterion cohort impact, and your population against the real one
Standing cohort, refreshed on each data cut	The same definition re-run, with a diff against the prior cut

SEQUENCE

Where we would start

01

Feasibility with an attrition funnel

The speed is the headline, but the funnel is what makes it credible. A number without the funnel is a faster guess.

02

Then target trial emulation

An external control arm is the highest-stakes instance of it, not a separate thing — one method serving both halves of your remit.

03

Before any of it, data fitness

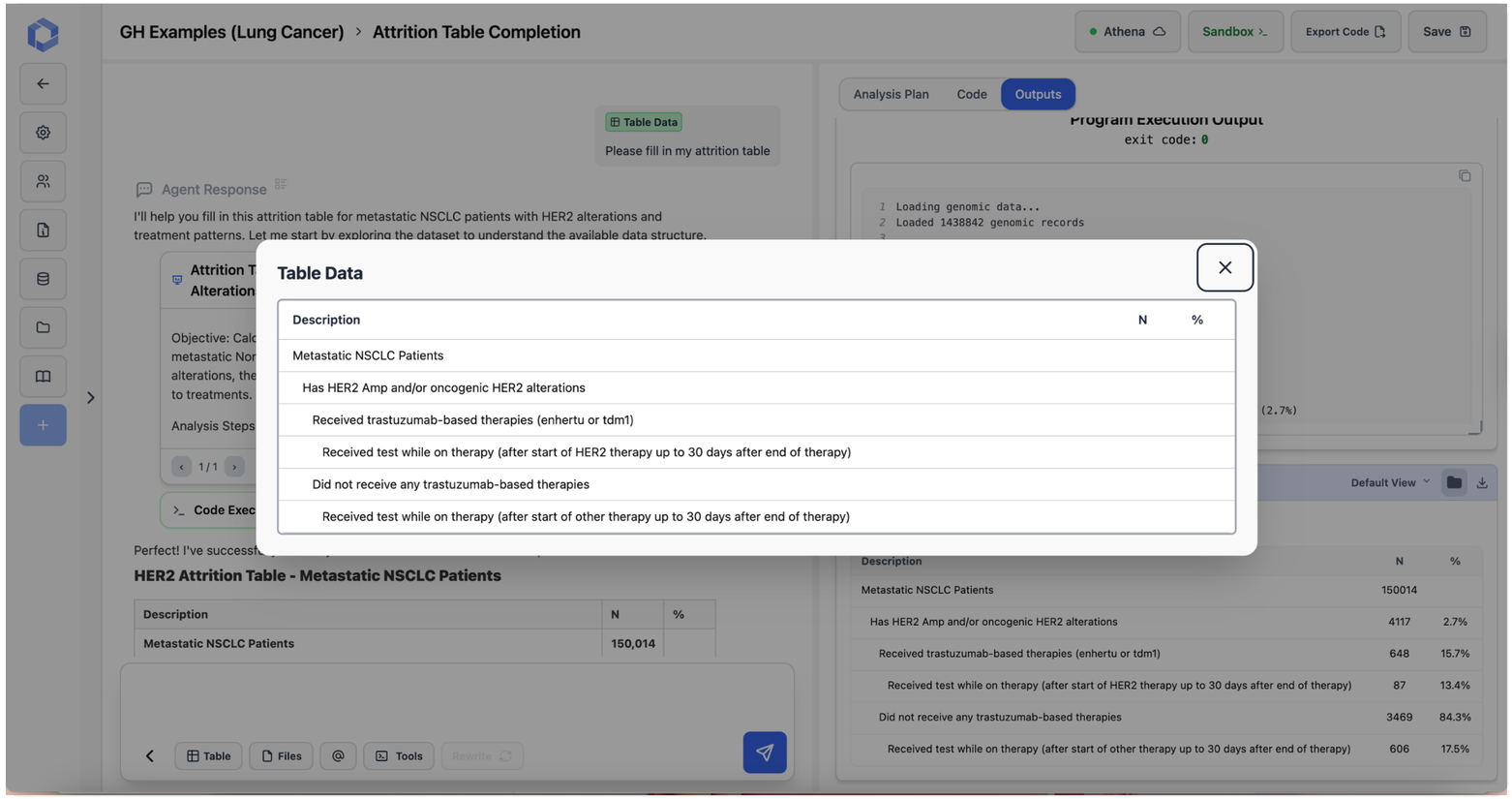
Whether your licensed sources can support the claim, which variables are proxied, and what limitation survives. Week one, not review.

ONE WORKFLOW, END TO END · STEP 1

Define the cohort and the question

Metastatic NSCLC with HER2 alterations, asked in plain language, with the analytical objective stated: treatment patterns, testing behavior, or survival.

The attrition table comes back with counts and percentages at every step, so the cohort can be inspected before anything is analyzed.

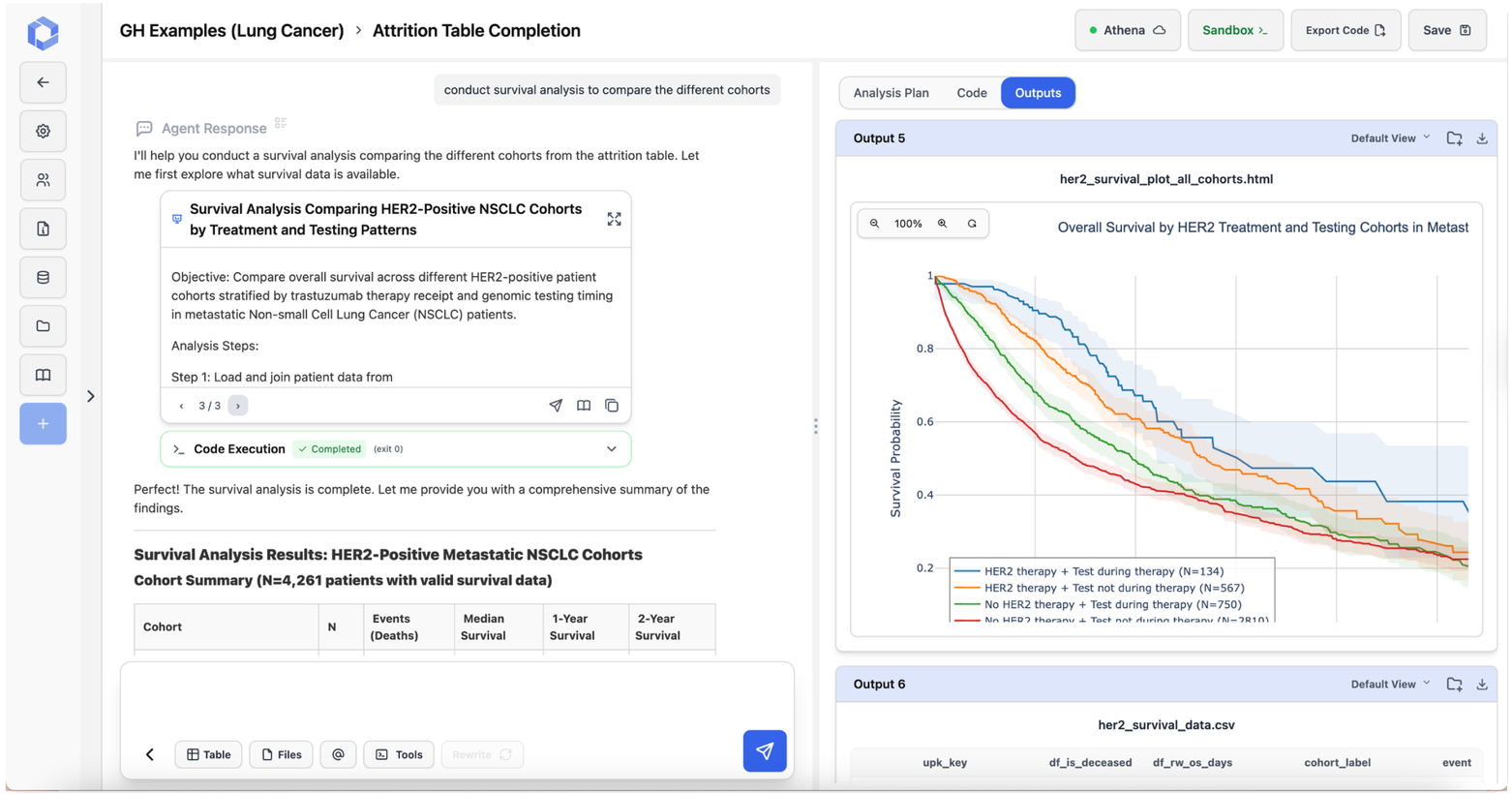


ONE WORKFLOW, END TO END · STEP 2

Run the survival analysis

Kaplan–Meier overall survival across the four cohorts, with consistent definitions of index date, censoring, and follow-up.

Plan, code, and output are all retained — the same session that defined the cohort produced the comparison.



02

Why this is defensible

What we would claim about ourselves, how the architecture holds, and what has been published.

CLAIMS

What we would say about ourselves

Eight of them, each checkable in the first week rather than taken on faith.

01 · WHERE IT RUNS

Data-neutral

No new data purchase, no internal decision about which vendor wins.

Runs where the data lives

Raw records stay in your environment.

Fits the estate in place

OMOP, claims, registry, EHR — alongside SAS, R, Python, Spotfire, Tableau.

02 · WHO DOES THE WORK

Your epidemiologists

Natural-language extraction puts the analysis with whoever has the question.

Bring your own code

Validated SQL, R, Python, and cohort definitions import as they are.

Programmable

Call it from your own scripts; save a study and re-run it on the next cut.

03 · WHAT HOLDS UP

Full audit trail

Reproducible definitions and analyses, suitable to support regulatory use.

Published methods

Defensible by citation rather than by assertion.

ARCHITECTURE

Secure by design, auditable by default

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, not a policy commitment.

Enterprise security reviews passed at Guardant Health and Regeneron. SOC 2 report, certificates, and completed vendor questionnaires on request.

SOC 2 Type II

ISO 27001

HIPAA · BAAs in place

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, whatever a prompt asks for.

Immutable audit trail with a human checkpoint: actions, data changes, and system events logged; 21 CFR Part 11 signature capability; encryption at rest and in transit; ALCOA+ throughout.

Formal computer-system validation is a separate track, and we would rather tell you where it stands.

METHODS

Models built on clinical research

Domain-specific models outperform general-purpose ones on clinical tasks, and the comparison is published rather than asserted.

LEADS

Literature mining trained on 21,335 systematic reviews and 453,625 citations; all-round improvement over GPT-4o on the paper's benchmarks.

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

1.3M+

clinical trial publications

850K+

protocols across 14 registries

150K+

patient notes

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

ENGAGEMENT

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

Deployed into your environment and operated by your epidemiologists — for functions that want to own and run their own tooling.

YOU OWN

The cohorts, the generated code, and the validated workflows built on it

1–3 months to fit the agents to your data model; integration plus licence

2

Delivered as a service

You buy the output, not software: a feasibility package, an external control arm, a comparative analysis — for teams under a date they cannot move.

YOU OWN

Protocol, cohort definition, executed code, results, and the audit trail

Priced per deliverable or milestone — an outcome, not a seat count

3

Build with us

For functions that have already decided to build. You have the engineers; what is missing is clinical development grounding.

YOU OWN

The pipelines, the skills, and the capability — knowledge transfer is the deliverable

Defined scope with a delivery date, never open-ended staffing

EVIDENCE

Relevant proof

Guardant Health

50+ data scientists and technical sales staff in production on oncology feasibility and real-world evidence.

3–5 days
feasibility, from 6–8 weeks

~3×
analyst throughput

40%
less external analytics

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, on analyses previously infeasible without dedicated infrastructure.

DSWizard

Nature Biomedical Engineering, 2026

A multi-step analysis that takes roughly three months by hand, completed in about a day — with a human setting the target and reviewing the results.

READING

Relevant publications

WHAT IT SUPPORTS	PAPER
Agent-run analysis you can check	DSWizard (Nature Biomedical Engineering, 2026) · Can LLMs Replace Data Scientists in Clinical Research? (2024)
Longitudinal EHR modeling	HALO (Nature Communications, 2023) · PromptEHR (EMNLP, 2022) · SynTEG (JAMIA, 2021)
Prediction on real-world populations	MediTab (IJCAI, 2024) · Evidence-driven spatiotemporal hospitalization prediction (Nature Communications, 2023)
Comparators without patient-level access	Individual patient data reconstruction from published trials (Machine Learning for Healthcare, 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, 40% average time saving

The full record is in the companion page, The Research Behind the Work — every workflow mapped to the papers underneath it, including where we have none to offer.

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

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

We show the emulated protocol and the diagnostics before anyone commits to a study — so you are judging a method.

CONTACT

contact@keiji.ai

Keiji AI