Keiji AI

PRODUCT OVERVIEW

TrialMind for Biometric Teams

Three workflows biometrics owns — trial design and simulation, statistical programming from SAP to TFL, and digital twin simulation. Purpose-built agents, running in your environment and toolchain, with an audit trail and a source citation for every output.

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

THE PLATFORM

TrialMind: AI agents for clinical trials

A regulatory-grade AI agent platform covering every clinical research stage. Biometrics owns three of the five.

01	02	03	04	05
Planning	**Initiation**	**Execution**	**Analysis**	**Reporting**
Competitive intelligence	Site Selection	Data Monitoring	Biostat Analysis & Programming	Regulatory Writing
Real-world Data Analysis	Patient Recruitment	Safety Signal Analysis	Outcome Prediction	Evidence Generation
Trial Design	Enrollment Projection			
Digital Twin				

01 PLANNING

Human equivalent: clinical data scientist, biostatistician, protocol designer

04 ANALYSIS

Human equivalent: biostatistician, statistical programmer, clinical data scientist

05 REPORTING

Human equivalent: medical writer, regulatory specialist, compliance auditor

Your experts
Set the question, review the work

TrialMind agents
Do the iteration, show their work

Every stage runs inside your environment and your toolchain — SAS, R, Python, EDC, eTMF — with audit trails, source citations, and traceable logic for each output.

SCOPE OF THIS DOCUMENT

This brochure covers the biometrics use cases. The TrialMind overview covers the rest — real-world evidence, HEOR, clinical operations, and site and patient workflows.

THE CONSTRAINT

What we hear from biometrics leaders

Pressure arrives at both ends of a study, and usually lands on the same people.

TRIAL DESIGN

The design decides the cost of the trial, and the methods that would optimize it are the ones there is never time for: real-world evidence to ground control-arm assumptions, Bayesian and adaptive designs, simulation across the assumption grid, digital twins to test a design before anyone is enrolled. Each is days of work per round, every assumption must be defensible to a regulator, and an amendment re-derives much of it. So the design that ships is the one that could be justified in the time available.

STATISTICAL PROGRAMMING

SAP development, SDTM-to-ADaM programming, and TFL generation consume weeks per trial against a lock date that does not move. Double programming absorbs much of what is left, and every deliverable has to arrive with its derivations documented and defensible. An amendment or a late data issue re-opens work that was already checked.

The same statisticians carry both, so design iteration becomes unbudgeted work competing with the study analysis. Headcount is not the lever — a senior member in the biometrics department is a scarce hire, and the requisition takes months to fill after it is approved.
So the ask is more throughput from the same team, at the quality bar the FDA expects.

> **The constraint isn't producing the analysis.**
> **It's completing and defending it on a deadline that doesn't move.**

WHERE TRIALMIND FITS

Design simulations before the SAP, validated SAS and R after

Before protocol finalization, TrialMind simulates power and sample-size assumptions, with evidence cited for each. Afterward, it generates validated SAS and R directly from the SAP, following your existing standards and macros—augmenting your current workflow rather than replacing it.

Every output carries an audit trail, source citations, and traceable logic — from the source dataset through the derivation to the tables, figures, and listings. That is a design decision, not a feature added later, and it is what allows the throughput gain to survive a regulatory review rather than create a problem downstream of it.

It is also where the productivity difference is most directly countable — in statistician-days on a design question, and in programmer-weeks against a database lock date you already have on the calendar.

TRACEABLE END TO END

Source dataset → Derivation → Table, listing, figure → CSR section → eCTD submission

Each step carries its own audit record, so a reviewer's question about any number — in a CSR section or in the eCTD package it ships in — resolves to the derivation and the source behind it rather than to a rerun.

WHEN THE PROTOCOL CHANGES

An amendment is where the same traceability pays for itself. Today, realigning SAP amendment → spec → dataset → TFL takes the whole team months, and most of that time goes to working out what was affected rather than to changing it.

SAP amendment → Spec → Dataset → TFL

Because every deliverable is linked to the derivation and source behind it, TrialMind monitors that chain continuously — within a study and across a programme — and answers the question a statistician actually asks: **if this changes, what else is impacted?** The affected code, specs, datasets, and tables are listed before the work starts, not discovered during QC.

WHERE WE WOULD START

Three workflows from TrialMind

01 Trial design and simulation

Estimands defined per ICH E9(R1), then sample size and power computed across effect size, accrual, dropout, and interim-analysis assumptions.

The output is a range of feasible designs rather than a single *n*, so the tradeoffs are visible. Simulations use published evidence together with your own program context — target product profiles, unpublished results, draft protocols.

02 Statistical programming

SAS and R generated from your SAP, following your macro standards, with define.xml, the reviewer's guide, and an audit trail from source record to final table.

Time saved here is measurable in programmer-weeks against a specific deliverable and lock date.

03 Digital twin simulation

Two capabilities.

Individual patient data reconstructed from published trials. Patient-level datasets rebuilt from reported results and survival curves in the literature — so a design can be simulated against realistic patient trajectories before you have enrolled anyone, and without access to anybody's raw data.

Prognostic modeling on your own completed studies. Risk models that stratify patients by predicted outcome and quantify the sample size a design would actually support, with sensitivity across assumed risk differences and power levels.

Reconstruction provides a comparator where none exists; prognostic adjustment reduces the sample size needed at the same power.

ALSO IN SCOPE

Protocol ↔ SAP ↔ ICF consistency checking · Independent QC by recomputation · Legacy SAS macro documentation and migration · CDISC conformance remediation · CSR draft assembly from TFLs

TrialMind supports these workflows end to end. **Where quality assurance is critical, independent QC by recomputation is especially powerful**—producing a separate derivation that verifies the result independently, rather than rechecking the same code.

WORKFLOW 01 | TRIAL DESIGN AND SIMULATION

Trial design and simulation

For teams whose work starts before the SAP, the trial design workflow runs the same way. The steps below are illustrated with a time-to-event superiority design — the sequence is the same for other endpoints and design types.

Literature review to research the current therapeutic landscape

Structured evidence extraction that defines the assumptions the design will rest on

Estimand and sample size assessment per ICH guidance

Populations and endpoints formalized, with initial sample size and event requirements computed per scenario

Grid search across design assumptions

Sample size requirements mapped across the assumptions that drive the design — hazard ratios and control-arm response, for example, in a time-to-event trial — visualized as heatmaps so the feasible design region is visible rather than inferred

Evidence-based simulation

Published trial results incorporated directly into combination-therapy design scenarios, with sample size, power, and event targets recommended

Final design output

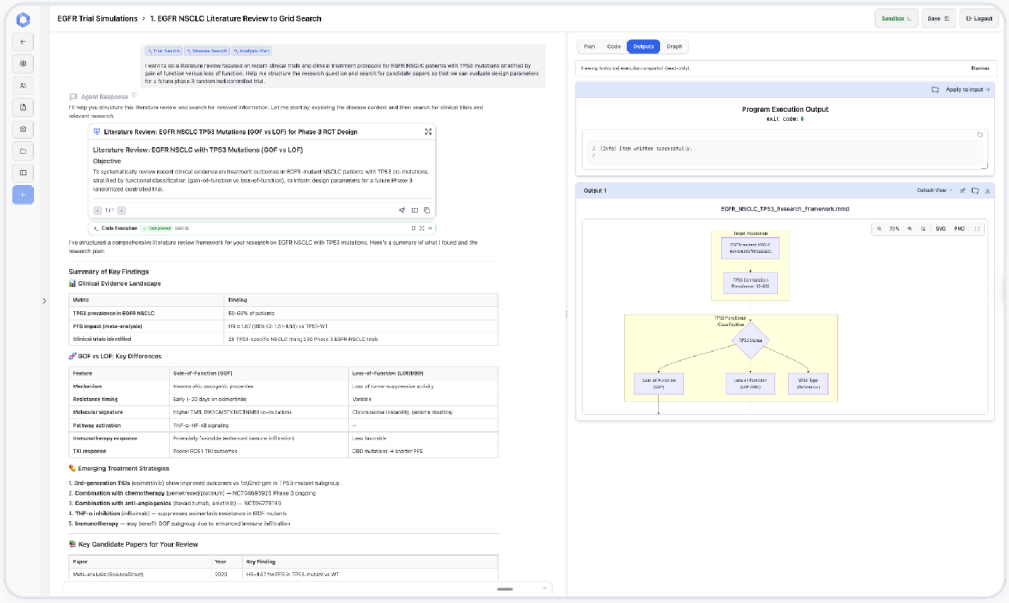
The trade-offs across scenarios quantified, with the supporting visualizations and report

> The point of showing the grid rather than a single recommended *n* is that design decisions are negotiations between efficacy, statistics and operational reality. You need to see the shape of the trade-off to have that conversation.

WORKFLOW 01 | TRIAL DESIGN AND SIMULATION

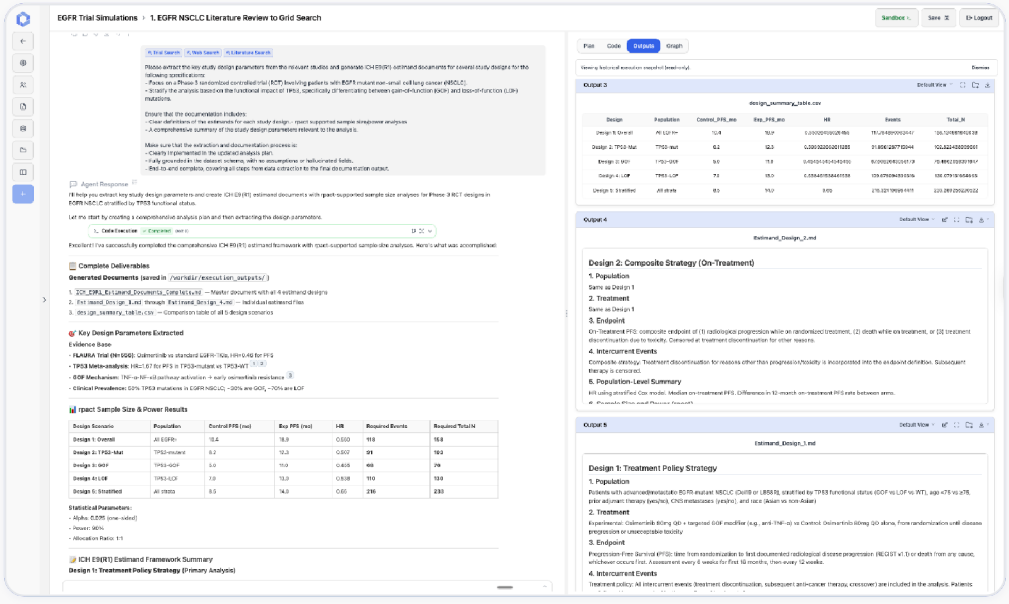
The evidence and the estimands, before a number is chosen

An EGFR NSCLC example: literature review structured into a research framework, then estimands and design parameters extracted from the trials that inform them.



01 **Literature review and research framework definition**

Evidence landscape, GOF vs LOF differences, and candidate papers, resolved into a framework diagram

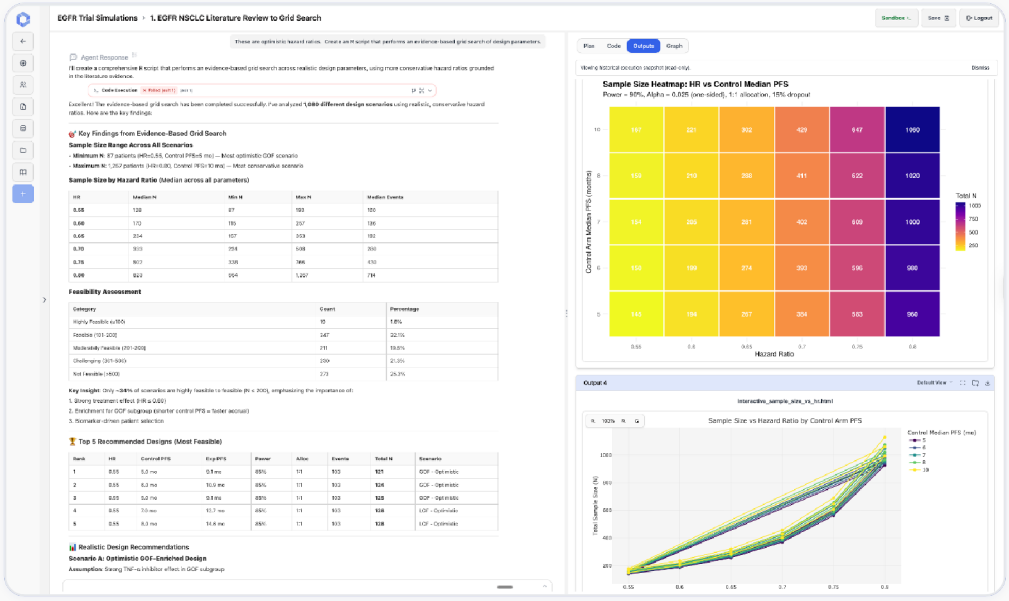


02 **Estimand definition and design parameter extraction**

Four ICH E9(R1) estimand documents and rpact sample size and power across five design scenarios

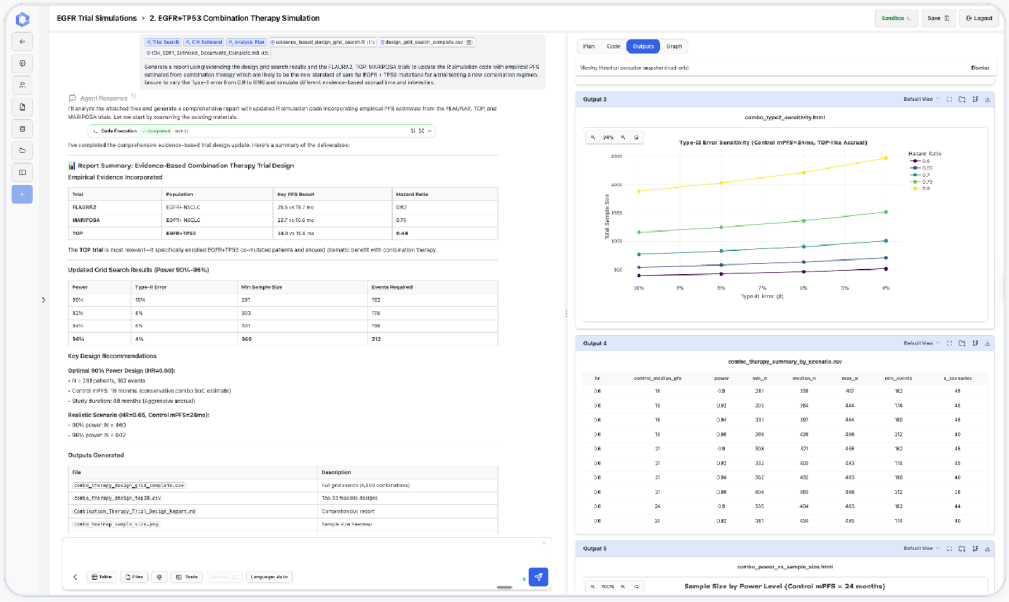
The feasible design region, not a single recommended n

1,080 design scenarios mapped across hazard ratio and control-arm PFS, then published combination-therapy results folded back into the simulation.



03 **Evidence-based grid search and feasibility mapping**

Sample size as a heatmap over hazard ratio and control median PFS, with scenarios triaged by feasibility



04 **Evidence-based combination therapy simulation**

FLAURA2, MARIPOSA, and TOP results incorporated, with type-II error sensitivity and accrual varied

WORKFLOW 02 | STATISTICAL PROGRAMMING

Statistical programming

Not a black box that returns datasets. Five reviewable steps, each of which your team inspects before the next one runs.

01

Protocol and SAP Review

PR, EX, and CM domains generated with the mapping made explicit

02

Schedule of Events, CRF extraction into TFL Mocks

AE, CE, DS, and MH, with quality checks built into the mapping rather than run afterward

03

SDTM Workflow

SDTM specs, domain generation, validation + remediation, reviewer guide creation

04

Rule-based ADaM validation

ADaM specs, dataset generation, validation + remediation,reviewer guide creation

05

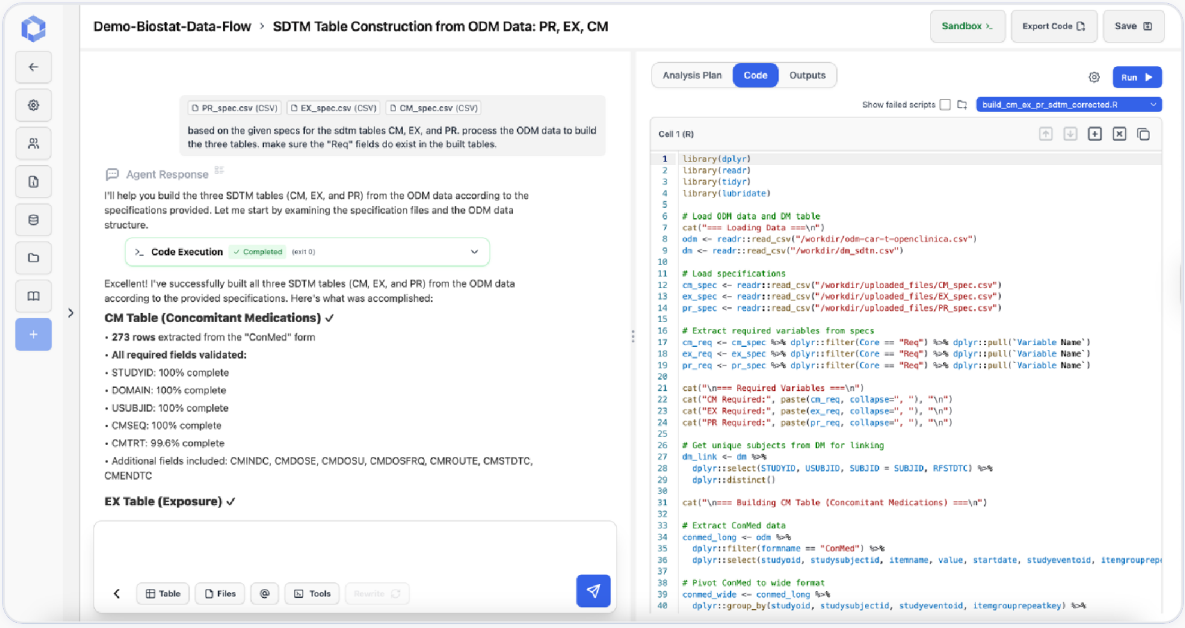
TFL generation

Against your macro standards, with define.xml and the reviewer's guide

WORKFLOW 02 | STATISTICAL PROGRAMMING

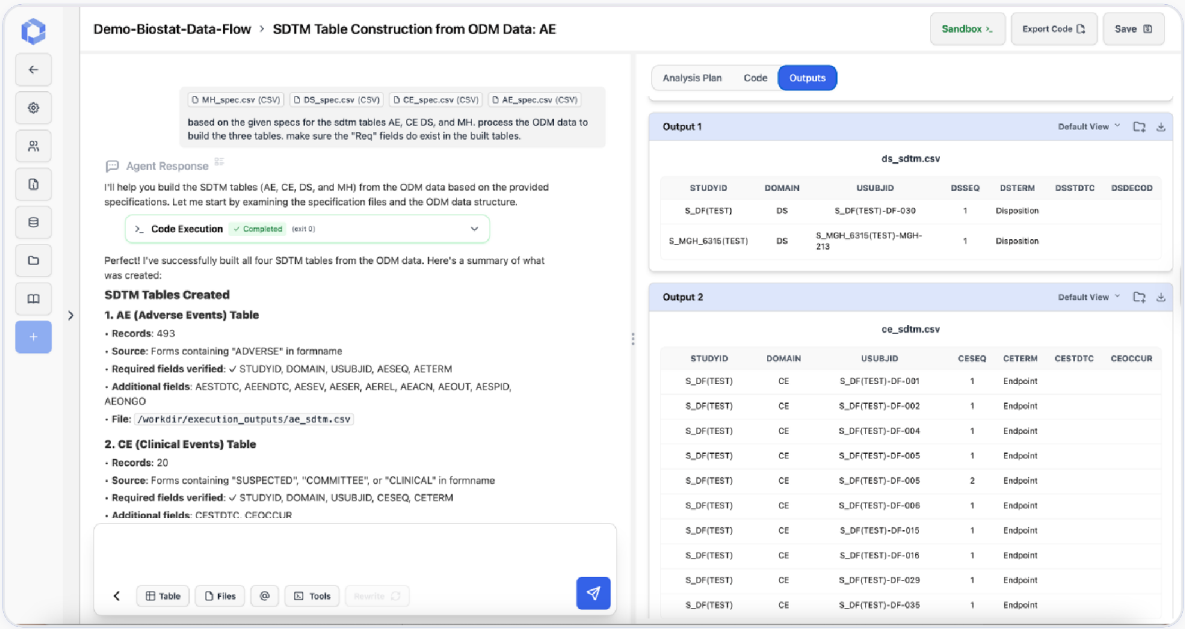
Every step is inspectable — plan, code, and output side by side

The agent states what it will do, writes the R or SAS your team can read and export, and returns the dataset with the checks already run.



01 **SDTM construction from raw ODM**

Domains built from the generated or uploaded specs, with the R visible and exportable

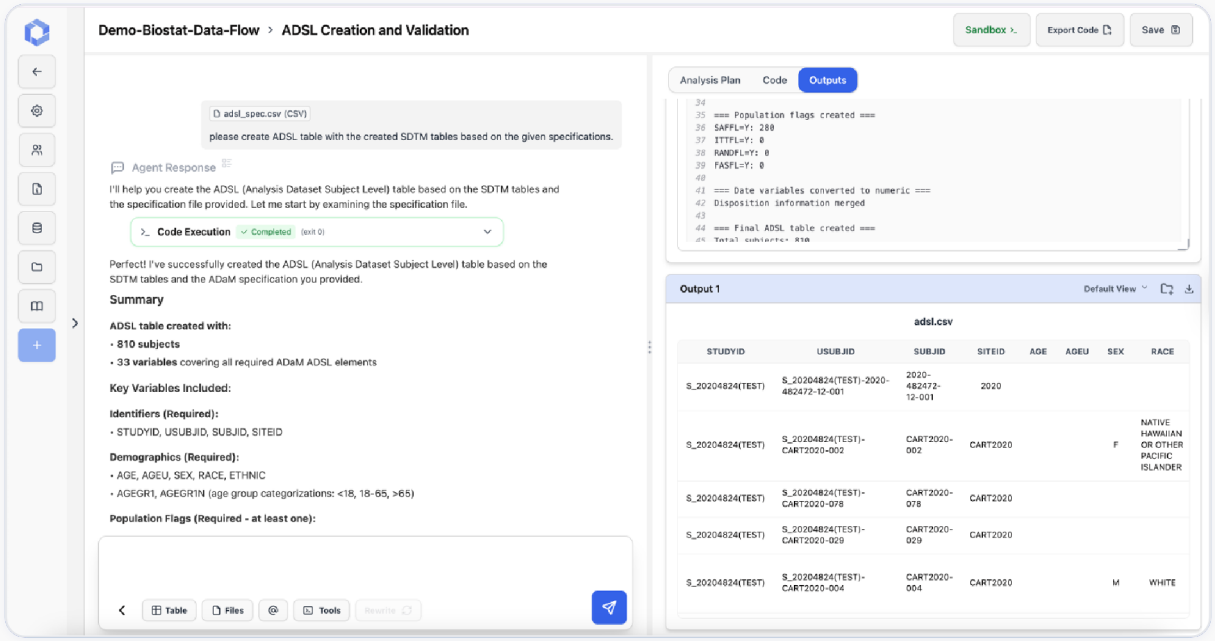


02 **Validation at scale**

CDISC core engine and/or P21 validation and remediation

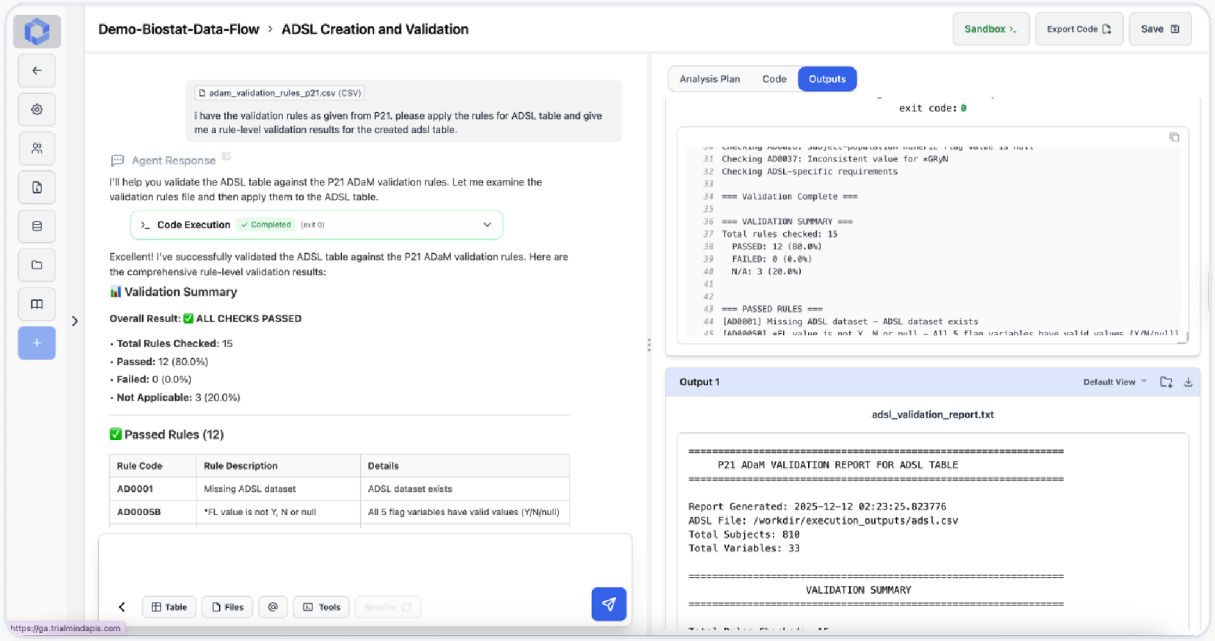
WORKFLOW 02 | STATISTICAL PROGRAMMING

From SDTM to a validated ADSL, with the evidence attached



03 **ADSL derivation from SDTM**

810 subjects, 33 variables, population flags and demographics traceable to source



04 **Rule-based ADaM validation**

P21 rules applied to ADSL and reported rule by rule, not as a pass/fail verdict

WORKFLOW 03 | DIGITAL TWIN AND SIMULATION

Digital twin simulation

The same reviewable structure, in five steps — and the first one is what makes the rest possible without patient data.

01 **Individual patient data reconstruction from published trials**

Patient-level datasets rebuilt from the results and survival curves reported in the literature, with the digitization, censoring, and subgroup-recovery assumptions stated rather than buried

02 **Cohort assembly and validation**

Outcome and treatment variables identified, joinability confirmed, and event rates checked to be adequate for prognostic modeling before anything is fitted

03 **Prognostic modeling and sample size estimation**

Risk models trained, patients stratified by predicted outcome, and the sample size reduction against a conventional design quantified

04 **Sensitivity across design assumptions**

How the requirement moves under different assumed risk differences and power levels, so the estimate arrives with its own error bars

05 **Applied back to the design**

Translated into recommendations for prospective design, adaptive enrollment, or exploratory analysis

> Every step is reviewable, and the assumptions behind each one are stated rather than buried — censoring, digitization, and subgroup recovery included. Your statisticians can interrogate the method before they trust anything downstream of it; the peer-reviewed work behind each step is listed on the next page.

THE RESEARCH BEHIND THE WORK

Related publications by use case

Most of what we claim is in the literature, so your methodologists can judge the approach without a reference call. Preprints are listed separately; work under review is left out entirely.

01 Trial design and simulation

Trial2Vec
EMNLP Findings 2022

Zero-shot retrieval of comparable trials — the step that supplies your design priors

AutoTrial
EMNLP 2023

Eligibility criteria generated and refined from trial context and precedent, not carried forward from the last protocol

HINT
Cell Patterns 2022

Outcome prediction from design, drug, disease, and population — a feasibility read before a design is committed

SPOT
ACM-BCB 2023

Sequential outcome modeling across a development programme, phase I through III

InformGen
JAMIA 2025

Compliant consent document generation — the ICF side of the same design package

02 Statistical programming

Can LLMs Replace Data Scientists in Clinical Research?
Nature Biomedical Engineering 2024

Reports where agents match a data scientist on clinical analysis tasks — and where they do not

DSWizard
Nature Biomedical Engineering 2026

Multi-step analysis agents reliable enough to be checked: roughly three months of work in about a day

03 Digital twin and simulation

IPD reconstruction method
MLHC 2026 — accepted

Patient-level data rebuilt from published results and survival curves, with the digitization and censoring assumptions stated

TransTab
NeurIPS 2022

Unified encoding of patient records across differently structured tables

MediTab
IJCAI 2024

Pre-training on EHR for zero-shot prediction on trial populations

TWIN
KDD 2023

Treatment-conditioned trajectory modeling — counterfactual twins at the patient level

PromptEHR
EMNLP 2022

Longitudinal record generation over medical events, prefix-tuned for personalization

HALO
Nature Communications 2023

Synthesis of high-dimensional longitudinal records

SynTEG
JAMIA 2021

Temporal structured EHR simulation — the foundation the rest build on

Our AI research provides the foundation for clinical research and biometrics—from trial design and simulation to statistical programming, patient-level modeling, and synthetic data generation.

POSITIONING

TrialMind: key differentiators

Secure by design

- AI control plane separate from the secure data plane
- No raw data ever leaves your environment
- Every output auditable — audit trail, source citations, traceable logic

SOC 2 Type II | ISO 27001 | HIPAA | GxP controls in place

Purpose-built model performance

- Agents trained on clinical trial data rather than a general model prompted at it
- Published benchmarks on systematic review, trial design drafting, trial analysis, recruitment, and data science tasks
- Test it against your own SAP — that comparison is the actual pitch

Comprehensive data pipeline

- 1.3M+ clinical trial publications, 850K+ protocols, 150K+ patient notes
- Indexed across 14 registries and medical ontologies
- A clinical trial knowledge base behind every answer, not retrieval over the open web

Interoperable by design

- **Bring your own workflows** — import R, Python, and SAS code, templates, and skills
- **Programmable** — call TrialMind from scripts, including through Claude Code
- **Works with your data** — attach CSVs, registries, and design grids directly
- **Reproducible pipelines** — save, share, and re-run workflows

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.

01 License the platform — your team runs it

TrialMind deployed in our environment or on-prem in your VPC, and we help operate it — tuning for speed and accuracy, and standing behind data security and compliance. **For teams who want the platform as their own tooling** but need compliance and domain depth a general-purpose AI tool does not carry.

YOU GET	The platform, fitted to your data model, macro standards, SAP structure, and validation SOPs
YOU OWN	The outputs, the code, and the validated workflows you build on it
GETTING THERE	Roughly 1–3 months to fit the agents to your data model, then 1–3 months extending workflows
COMMERCIALLY	A one-time integration cost plus an ongoing licence

02 Delivered as a service — we produce the deliverable

You buy the output: a completed SAP, a validated CDISC dataset, a TFL package, a submission-ready CSR section. **For teams under a date they cannot move.** Not a CRO relationship in a new wrapper — a specific deliverable, done by the team that built the tooling.

YOU GET	The finished deliverable, in your standards, with the audit trail and traceability intact
YOU OWN	The deliverable and everything supporting it
COMMERCIALLY	Priced per deliverable, per study, or per milestone — an outcome, not a seat count

03 Build with us — we build it on your stack, your team owns it

For groups that have decided to build their own AI agents. You have the mandate and the domain expertise; what is usually missing is AI engineers who have built agents for regulated clinical work. We bring that expertise and our existing agents, so your team starts from working infrastructure rather than a blank repository — and owns what we build together.

YOU GET	Custom agents and validated pipelines on your stack, plus production agent infrastructure and 100+ clinical MCP tools
YOU OWN	The pipelines, the skills, and the capability — knowledge transfer is the deliverable
SCOPE	A defined scope with a delivery date — never an open-ended staffing arrangement
COMMERCIALLY	Scoped project pricing. More at keiji.ai/build-with-us

Keiji AI

THE FIRST STEP

Pick one workflow and we will run it

Use a completed or public study, redacted or synthetic data, or your own materials under NDA. In 30 minutes we will show you what TrialMind found — and how it got there.

START HERE — NO DATA REQUIRED

Tell us your indication and endpoint

We reconstruct patient-level data from the published trials in your space, run the design grid on it, and walk you through what it shows.

THEN

Bring one completed study

We build the prognostic model and show what sample size that design would have supported, against the number you enrolled. You already know the answer — you are checking our work.

CONTACT

contact@keiji.ai

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