CASE STUDY · EXACT— CLINICAL TRIAL MATCHING INFRASTRUCTURE

EXACT Matching, Explained

How HealthKey’s EXACT (EXtracting Attributes from Clinical Trials)engine turns a structured patient record into transparent clinical trial eligibility — and how any organization can build on it

The trial-matching capability behind CancerBot, available as infrastructure

Powered by CTOMOP · EXACT · Open source on GitHub

266	**3**	**98%**	**0**
ELIGIBILITY-READY FIELDS	VERDICT STATES PER CRITERION	OF PATIENTS MISS TRIALS TODAY	BLACK-BOX SCORES

Executive Summary

Clinical trial matching is widely treated as a search problem. It is not. It is a data problem. Trials are not hard to list — they are hard to match, because matching requires a patient record complete and structured enough to evaluate against eligibility criteria written in dense clinical language.

HealthKey solves that problem with two open-source properties working together. CTOMOP is a clinical-trial-ready superset of the OMOP Common Data Model that turns a messy, multi-provider history into 266 eligibility-ready fields. EXACT clinical trial matching system is a stateless matching engine that consumes that structured profile and returns, for every trial, a per-criterion verdict — passed, failed, or indeterminate — with the exact patient value and trial threshold behind each decision.

This capability was first proven in production by **CancerBot**, an AI-powered trial-matching service for cancer patients built entirely on CTOMOP and EXACT. CancerBot demonstrated that the model works for real patients across multiple myeloma, follicular lymphoma, and breast cancer. This case study explains how EXACT (EXtracting Attributes from Clinical Trials) works, what made CancerBot succeed, and — most importantly — how any organization can use the same infrastructure to deliver trustworthy trial matching of its own.

> *“If your record told you that you don’t qualify for a trial, you should be able to see exactly which field and which threshold produced that answer.” — Paul Ahlstrom*
> *HealthKey CEO, on why EXACT is built to explain itself*

The Problem: Trial Matching Is Broken at the Data Layer

2% enrollment, and a 98% blind spot

Historically, only about 2% of cancer patients ever enroll in a clinical trial — not because suitable trials do not exist, but because finding them is confusing, time-consuming, and opaque. The remaining 98% never find out what they might have qualified for.

The failure is not a shortage of trials. It is a chain of breakdowns between the patient and the trials that already exist:

- **Criteria are written for clinicians, not patients.** Eligibility language is dense, technical, and assumes detailed lab values most patients have never been told how to gather.
- **Trial information is fragmented.** Open trials are spread across many registries, each with its own structure and update cadence.
- **Patient data is incomplete.** Industry data shows roughly 40% of clinical trial delays are caused by incomplete patient data, and a complete longitudinal record is something 87% of patients simply cannot access.

- **Existing tools return scores, not reasons.** A patient told only that they are a ‘62% match’ learns nothing actionable. Neither does the navigator trying to help them, or the oncologist asked to sign off.

The cost lands on everyone at once. Patients who could benefit from cutting-edge treatments never hear about them. Trial sponsors struggle to hit enrollment targets. And organizations that try to build matching themselves discover that the hard part was never the user interface — it was producing a patient record trustworthy enough to match against, and a verdict trustworthy enough to act on.

> *2% of cancer patients enroll in clinical trials. 98% never find out what they qualify for.*

The Solution: A Structured Record Meets an Explainable Engine

HealthKey's answer is to fix both halves of the problem in the right order: first make the patient record structured enough to reason over, then match it with an engine that shows its work. Two open-source properties carry that load.

CTOMOP — the structured patient record

CTOMOP is a clinical-trial-ready superset of OMOP CDM v6.0. It extends the standard model with oncology and genomics tables, a richer episode model for lines of therapy, and a denormalized **PatientInfo** projection of 266 eligibility-ready fields. Critically, every computed field — therapy-line count, refractory status, measurable disease per IMWG, TP53 disruption, lymphocyte doubling time — is derived **once, in one place, with auditable rules.** There are no ad-hoc mappings and no one-off fields. Every lab value, therapy line, and biomarker is stored the same way for every patient.

EXACT — the explainable matching engine

EXACT is a stateless clinical trial matching engine. (EXACT stands for EXtracting Attributes from Clinical Trials) It consumes a CTOMOP-aligned patient profile and a structured trials catalog, then evaluates every eligibility criterion of every trial. For each criterion it returns one of three verdicts — **passed**, **failed**, or **indeterminate** — together with the specific patient field value and the specific trial threshold that produced it. Nothing is persisted inside the matcher. Nothing is hidden inside a score.

Why three verdicts, not a score

The third verdict is the one that matters most. A binary match/no-match system has to guess when data is missing, and a guess is indistinguishable from a fact once it becomes a number. EXACT instead names the gap: an indeterminate verdict tells a patient or navigator exactly which missing value stands between them and a definitive answer — turning an unknown into a concrete next step, often a single lab test away.

Illustrative EXACT trace for a single trial

Eligibility criterion	Verdict	Patient value vs. trial threshold
Confirmed diagnosis	**PASSED**	Multiple myeloma — meets required diagnosis
Prior lines of therapy	**PASSED**	Patient: 2 lines · trial requires ≥ 1
Measurable disease (IMWG)	**PASSED**	M-protein present — measurable criterion met
ECOG performance status	**FAILED**	Patient: ECOG 3 · trial requires ≤ 2
Cardiac ejection fraction	**INDETERMINATE**	No recent echocardiogram on record

Every verdict references a real value and a real threshold. The patient and navigator act on reasons — including knowing which one test would resolve the open question.

How EXACT Works, End to End

For an organization integrating EXACT, the flow from raw records to an explained match is consistent regardless of the front-end experience built on top.

1. Assemble the CTOMOP record

Patient data enters as FHIR R4 bundles from EHRs, hospital portals, and lab systems — or through structured intake questions. HealthKey’s ETL pipeline, using OHDSI tooling alongside its own PHRofile and PHRogram components, maps every clinical concept to standard OMOP vocabularies and produces a clean CTOMOP record. The PHResolution step surfaces conflicting values for review so errors are caught before they ever reach the matcher.

2. Derive the PatientInfo projection

From the CTOMOP record, the 266-field PatientInfo projection is computed. This is where complex, multi-table clinical logic — inferring lines of therapy, staging, refractory status, biomarker presence — is resolved once into clean, eligibility-ready fields. The hard clinical reasoning happens here, not inside the matcher.

3. Run EXACT against the trials catalog

PatientInfo is handed to EXACT along with a structured, up-to-date trials catalog. EXACT evaluates each trial criterion-by-criterion and returns the per-criterion trace. Because the engine is stateless, the same inputs always produce the same explained output — a property that matters enormously for auditing and for clinical trust.

4. Present reasons, with a human in the loop

Results reach the patient in plain language: which trials match, and precisely why. A patient navigator reviews the matches, answers questions, and confirms understanding before any outreach. EXACT informs the decision; it does not replace the people making it.

The pipeline at a glance

Stage	What happens	Output
Ingest	FHIR bundles and intake data captured from every source	Raw multi-provider data
Transform	OMOP vocabulary mapping; PHRofile / PHRogram ETL	Clean CTOMOP record
Reconcile	PHResolution flags conflicts for human review	Trusted CTOMOP record
Project	266 computed eligibility fields derived once	PatientInfo profile
Match	EXACT evaluates every criterion of every trial	Per-criterion verdict trace

Stage	What happens	Output
Review	Navigator confirms matches with the patient	Actionable, explained matches

Proof in Production: CancerBot

EXACT is not a prototype. It runs in production today inside CancerBot — the first AI-powered clinical trial matching service built on an open, oncology-grade patient record, connecting cancer patients with potentially life-saving trials in minutes rather than months.

Built by a patient, for patients

CancerBot's founder, Adam Blum, was diagnosed with follicular lymphoma and experienced firsthand how broken the trial-finding process was. He chose to build CancerBot on HealthKey's open infrastructure for a specific reason: so that the patient data model and the matching engine could be audited, extended, and trusted rather than taken on faith.

What CancerBot demonstrates

One data standard. Because CTOMOP's OMOP-aligned schema stores every lab value, therapy line, and biomarker consistently, CancerBot never has to maintain ad-hoc mappings or one-off fields.

- **Explainable by design.** Every eligibility verdict CancerBot shows references the exact patient value and trial threshold — so patients and navigators act on reasons, not opaque scores.
- **Open-source foundation.** Because CTOMOP and EXACT are published on GitHub, CancerBot's matching logic can be audited by patients, providers, and researchers — building the trust that clinical trial enrollment requires.
- **Multi-cancer from the start.** CancerBot supports multiple myeloma, follicular lymphoma, and breast cancer — each with its own disease-specific computed fields, all on the same engine.

Disease coverage proven on the same engine

Cancer type	Disease-specific fields EXACT evaluates
Multiple Myeloma	Lines-of-therapy inference, ISS staging, M-protein, FISH cytogenetics
Follicular Lymphoma	FLIPI score, GELF criteria, histologic grade, transformation status
Breast Cancer	HR / HER2 status, BRCA, TNM staging, prior therapy history

The pattern is the point: adding a disease means contributing computed fields to CTOMOP, not rebuilding the matcher. The engine that matched a myeloma patient yesterday matches a breast cancer patient today.

> *CancerBot connects cancer patients with potentially life-saving trials in minutes, not months — built entirely on CTOMOP and EXACT.*

How Organizations Build on EXACT

CancerBot is one application of EXACT. The same infrastructure is available to any organization that needs trustworthy trial matching — and the open-source foundation means building on it does not mean betting on a vendor's black box.

Who EXACT serves

Organization	What EXACT enables
Patient nonprofits & advocacy groups	Offer members transparent, disease-specific trial matching without building a clinical data team — the model HealthTree Foundation follows for Cure Hub.
Pharmaceutical & biotech sponsors	Reach eligible patients faster and reduce enrollment delays caused by incomplete data, with an auditable record of why each patient matched.
Hospitals & cancer centers	Give patients pre-computed trial eligibility and give clinicians a verdict trace they can inspect rather than a score they must trust blindly.
Health-tech builders	Embed production-grade matching via CTOMOP + EXACT instead of reinventing oncology data modeling and eligibility logic from scratch.
Researchers	Extend CTOMOP to new diseases by contributing computed fields, on an OMOP-standard schema compatible with the wider OHDSI ecosystem.

Why integration is low-risk

Open and auditable. Both CTOMOP and EXACT are published under permissive licenses on GitHub. Clinicians can audit the eligibility rules; researchers can extend the model; engineers can read exactly how a verdict is reached.

- **Standards-based, not proprietary.** CTOMOP is a superset of OMOP CDM — the global standard for observational research — so an organization's data stays compatible with the entire OHDSI tooling ecosystem.
- **Stateless and reproducible.** EXACT persists nothing and is deterministic: identical inputs yield identical, explainable outputs, which makes validation and clinical sign-off straightforward.
- **Secure by foundation.** HealthKey is HIPAA compliant, SOC 2 Type II audited, and GDPR ready, with end-to-end encryption, patient-controlled access, configurable data residency, and a strict no-data-selling commitment.

The division of labor

The HealthTree Foundation partnership is the clearest template. HealthTree owns the trust and community layer — the patient relationship, education, and support. HealthKey provides the infrastructure layer that ingests fragmented records, standardizes them, and powers matching.

An organization brings its patients and its mission; EXACT brings the engine. Neither has to become the other.

> *The patient model is open. The matching engine is open. The infrastructure built on top is what HealthKey provides.*

Why Explainability Is the Whole Point

It would be easier to ship a single match percentage. HealthKey deliberately does not, and that choice is the core of the product.

Clinicians can audit it. The logic that decides whether a patient is eligible for a trial — or refractory to a therapy — should be inspectable by a human oncologist, not buried inside a vendor binary. Every derivation rule in CTOMOP is in the open.

Patients can verify it. When a patient is told they do not qualify, EXACT can show the exact field and threshold behind that answer. The verdict is legible; the code makes it accountable.

Navigators can act on it. An indeterminate verdict is a to-do list. It points to the specific missing value — a lab, an imaging result — that would convert an unknown into a definitive yes or no. Reasons create next steps; scores do not.

In a domain where a wrong answer can cost a patient a treatment option, trust cannot be asserted — it has to be demonstrable. Explainability is how EXACT demonstrates it, every single verdict.

Key Takeaways

- **Trial matching is a data problem.** The hard part is a complete, structured patient record — not search. CTOMOP solves that with 266 eligibility-ready fields derived once, with auditable rules.
- **EXACT matches by reasons, not scores.** Every criterion returns passed, failed, or indeterminate, with the exact patient value and trial threshold behind it.
- **Indeterminate is a feature.** Naming the missing value turns an unknown into a concrete, often single-test next step for the patient.
- **Proven in production.** CancerBot runs on CTOMOP and EXACT today across three cancer types — evidence the model works for real patients.
- **Available as infrastructure.** Nonprofits, sponsors, providers, and builders can deliver trustworthy matching without rebuilding oncology data modeling — the HealthTree partnership is the template.
- **Open source lowers the risk.** Because CTOMOP and EXACT are on GitHub and built on OMOP, organizations adopt auditable standards rather than a proprietary black box.

Learn More & Get Started

[HealthKey — Product, security, and case studies]()
[CTOMOP on GitHub — the open clinical-trial-ready patient model]()
[EXACT on GitHub — the open explainable matching engine]()
[CancerBot — EXACT trial matching in production]()

HealthKey · HIPAA Compliant · SOC 2 Type II · GDPR Ready · Built on OMOP, powered by FHIR