The History of Clinical Trial Recruitment (Prior to AI)

Douglas Bell, MD PhD
Co-Director, UCLA CTSI Informatics Program
Program Director, UCLA Clinical Informatics Fellowship
Professor, Department of Medicine, DGSOM

Early Clinical Trial Milestones

- 562 BC: Book of Daniel
  - Israelite prisoners in Persia didn't want to eat gentile meat, some dying
  - Given diet of beans, vegetables; after 10 days, healthier c/w other captives
- 1747: James Lind, British sailors
  - 12 in sick bay with scurvy, given 1 of 6 treatments e.g. 2 spoons vinegar tid
  - Pair who received oranges recovered in 6 days
- Placebo control 1st used 1863, safety outcome 1937, double-blind 1944
- 1947: British RCT of streptomycin for TB treatment
  - Patients recruited from TB clinic

Clinical Trial Recruitment

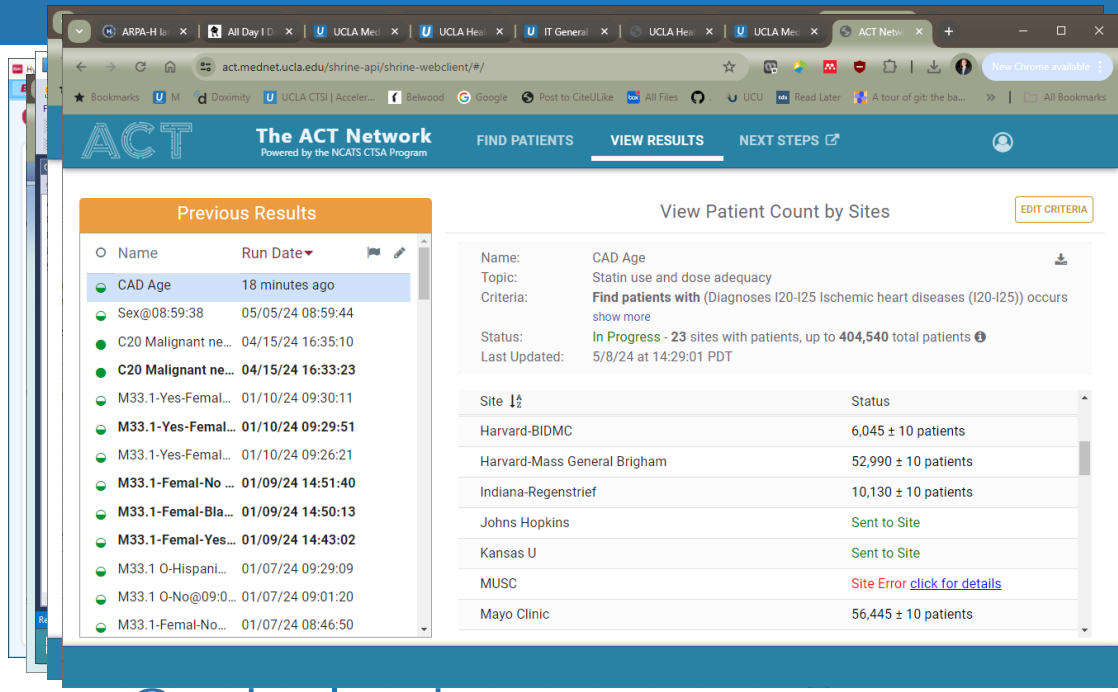
- Recruitment is fundamental for clinical trials
  - Validity: Was pre-planned sample size achieved?
  - Generalizability: Was population representative?
- The many options for contacting patients fall in 2 categories
  - Healthcare-based outreach
    - e.g. Specialty clinic, Doctor referral, EHR search, Patient registries
  - Public outreach
    - e.g. Newspaper ads, Health fairs, Social media, ResearchMatch
- Challenges
  - Messages often ignored, perception of burden, reluctance to be assigned

Strategies for Improving Recruitment

- 2010 Cochrane Review of 37 studies
  - Increasing awareness of the health problem (RR 1.48)
  - Video about the health condition (RR 1.75)
  - Monetary incentives (RR 1.39 – 1.53)
  - Not effective: recruiter differences, increasing patients' understanding of the trial process, varying consent design, varying method of randomization
- 2018 update: 68 studies
  - Open label rather than blinded, placebo (**+10%**)
  - Telephone reminders to people who do not respond to a postal invite (**+6%**)
  - Moderate-certainty evidence for 8 others (of 72)
    - Including: financial incentives, participant information, recruiter information

EHR-based search for eligible patients

- Self-service tools
  - Slicer-dicer (Epic)
  - I2B2 (UCLA only)
  - DDR (UCLA only)
  - LADR, ENACT (SHRINE)
- Mediated
  - PCORnet
- Getting access: i2b2, ladr or act@ctsi.ucla.edu



Health care outreach options

- Person making contact
  - Patient's own or other provider vs. RA or research personnel, others
- Timing of contact
  - During a health care encounter vs. asynchronous
- Medium of contact
  - Print materials (clinic or mailed), telephone, portal messages, email

One Example: TOPAZ Study

- RCT of IV zolendronate to prevent hip fracture among patients living with Parkinson's disease, 2021-24
  - Recruitment in the movement-disorders clinic
    - 35 patients eligible, consented and randomized
  - MyChart message to eligible patients
    - MyChart invitation sent to 523 eligible patients (with no visits in movement-disorders) + 3-month follow-up message to non-responders
    - 4 patients eligible, consented and randomized
      - 11 others enrolled but dropped out before randomization

Visioning the future of clinical trials recruitment and trial matching using AI

Arash Naeim, MD PhD
Co-Director, Center for AI & SMART Health
Co-Director, Clinical and Translational Science Institute
Chief Medical Officer for Clinical Research
Associate Dean for Clinical Research

Top Down: CT.GOV or CTMS (Trial Registry) First Strategy

1. Starts with structured eligibility criteria (e.g., from ClinicalTrials.gov or similar databases).

2. AI, especially Natural Language Processing (NLP), is used to analyze unstructured data in patient electronic health records (EHRs), such as clinician notes, diagnoses, or test results.

3. The aim is to automatically match patients to trials by comparing their characteristics (extracted from EHRs) against the trial requirements, essentially filtering eligible patients efficiently.

Jin, Q., Wang, Z., Floudas, C.S. *et al.* Matching patients to clinical trials with large language models. *Nat Commun* **15**, 9074 (2024). https://doi.org/10.1038/s41467-024-53081-z

Top Down: Advantages and Disadvantages

- **Advantages**:
  - **Scalability**: ClinicalTrials.gov provides a centralized and comprehensive database of trials, making it easier to build a pipeline that scales across many trials.
  - **Data Quality**: Trial criteria are often more structured and standardized compared to raw EHR notes, reducing preprocessing burdens.
  - **Fewer Privacy Concerns**: Working primarily with public trial data avoids some privacy and compliance issues associated with patient data.

Top Down: Advantages and Disadvantages

- **Disadvantages**:
  - **Contextual Gaps**: Patient-specific nuances in EHRs (e.g., history or detailed clinical notes) may not be captured as effectively.
  - **Domain-Specific Limitations**: General-purpose models may struggle with domain-specific eligibility criteria, whereas domain-specific models might need tailored training data for parsing non-oncology trials.

Bottom Up: EHR First Strategy

1. This approach focuses on leveraging AI to create a structured understanding of concepts within EHRs that are relevant to clinical trials.

2. The AI models identify and define clinical concepts (e.g., disease states, biomarkers, medication histories) from raw EHR data.

3. These structured concepts are then used to match patients to relevant trials, bypassing the need for explicitly mapping criteria and instead using predefined "concept libraries" or frameworks derived from EHR data.

Gupta, S., Basu, A., Nievas, M. *et al.* PRISM: Patient Records Interpretation for Semantic clinical trial Matching system using large language models. *npj Digit. Med.* **7**, 305 (2024). https://doi.org/10.1038/s41746-024-01274-7

Bottom Up: Advantages and Disadvantages

- **Advantages**:
  - **Personalized and Comprehensive Matching:** Uses granular patient data from EHRs, including medical history, demographics, and biometrics, allowing for precise matching to trial requirements. **Inclusion of Hard-to-Reach Populations:** Helps include diverse patients who might be missed in top-down approaches, especially those with complex medical histories.
  - **Real-Time Data Utilization:** Enables the use of the latest patient information, making it dynamic and reflective of current health conditions.
  - **Improved Patient-Centricity:** Focuses on patient needs and data rather than starting with trial sponsor-driven criteria, increasing the likelihood of finding trials that truly fit the patient.

Bottom Up: Advantages and Disadvantages

- **Disadvantages**:
  - **High Data Quality Requirements:** Relies on EHRs, which often have incomplete, unstructured, or inconsistent data, requiring significant preprocessing and validation.
  - **Scalability Issues:** Extracting and analyzing patient-level data across institutions or regions can be resource-intensive and challenging to scale.
  - **Lack of Standardization:** EHR systems vary widely in data formats and quality, which complicates interoperability and data aggregation.
  - **Limited Trial Awareness:** Without starting with eligibility criteria or trial availability, it may not efficiently narrow down the scope to active trials, potentially increasing false-positive matches.

Hybrid Model

1. This approach focuses on leveraging structured trial eligibility criteria guide AI-driven extraction and standardization of patient data from EHRs.
2. Trial criteria inform what data to prioritize in EHRs.
3. AI models then standardize, enrich, and match patient data to trials, ensuring scalability and accuracy.

Shriver SP, Arafat W, Potteiger C, et al. Feasibility of institution-agnostic, EHR-integrated regional clinical trial matching. *Cancer*. 2024; 130(1): 60-67. doi:10.1002/cncr.35022

Hybrid Model: Advantages and Disadvantages

- **Advantages**:
  - **Broader Coverage**: Combining top-down criteria alignment with bottom-up data extraction ensures comprehensive matching for both local and regional trials.
  - **Scalability**: Allows for a regional search framework that can adapt to varying population densities, benefiting both urban and rural areas.
  - **Reduced Resource Requirements**: The use of core variables minimizes the burden on EHR systems and providers while maintaining effectiveness.

Hybrid Model: Advantages and Disadvantages

- **Disadvantages**:
  - **Data Quality in EHRs**: Low fill rates for critical variables like biomarker status (6.2%) and metastasis indication (24%) hinder automation and require manual augmentation.
  - **Trial Criteria Ambiguities**: Non-public trial criteria and complex eligibility requirements may reduce the reliability of screen-positive matches.

Future AI Solution: Missing Data Problem

- Cancer Stage
- Biomarker Data
- Genomic Data
- Previous Lines of Treatment
- Biometric Data

Future AI Solution: Flexibility in LLM selection

- **ChatGPT-4**:
  - **Strengths:**
    - Vast training data and language understanding make it highly effective at interpreting the unstructured text and can adapt to diverse criteria across many specialties and trial
  - **Challenges**:
    - Needs fine-tuning to ensure optimal extraction and standardization of highly specific medical terminologies with potential inefficiencies in domains requiring deep expertise, like oncology-specific eligibility criteria.
- **Domain-Specific LLM**
  - **Strengths**:
    - If tailored can excel at parsing and structuring criteria for domain-related trials, offering higher precision for this domain.
  - **Challenges**:
    - Limited applicability outside its domain focus. May need complementary models or additional infrastructure for trials beyond its specialization.

Future AI Solution: AI Costing & Efficiency

$$\text{running time} = \left(\frac{\#\text{ of input tokens}}{\text{input speed}*3600} + \frac{\#\text{ of output tokens}}{\text{output speed}*3600} \right) \text{hours}$$

Future AI Solution: Beyond the walls of the institution

1. **Decentralized clinical trials and direct to consumer recruitment**

2. **Multi-institutional or regional recruitment across multiple hospitals and AMCs with potentially different EHRs**

Future AI Solution: Study, Sponsor, or Institution Locus of Control and Driver

1. Will AI solutions be available on a study-specific basis?

2. Will Sponsor's invest in infrastructure and expect that all their sponsored studies will use it?

3. Do institutions invest to ensure a common approach across all trials?

Future AI Solution: Features and Functionality

- Content Available
  - CT.gov listing
  - Protocol
  - Consent
  - Coverage Analysis
  - Medicare Coding
  - Budget
  - Insurance claims
- AI for Study Start-Up
  - Facilitation of CT.gov submission
  - Protocol and consent creation
  - FDA IND and IDE document preparation

Future AI Solution: Features and Functionality

- Recruitment
  - Social Media Campaigns
  - Navigation Chat Bots
  - Predicting which patient are more likely to participate in a clinical trial.
    - EHR Data
    - Behavioral Data Integration

- Addressing Disparities
  - **Equity Based Approaches**: Identify patient with barrier to participation.
  - **Personalized Recruitment Messaging:** Personalizing outreach to highlight trial and community value to underserved populations.

Future AI Solution: Features and Functionality

- Time Commitment for Participation
  - Estimating the time commitment for participation.
  - AI models to extract information from protocols (e.g., visit schedules, procedures, and follow-ups) to estimate/simulate the total time commitment for participant on the trial.
- Cost of Participation
  - Predict out-of-pocket expenses
  - Geospatial Analysis for travel and distance costs
  - EHR and Claims Data Integration to estimate **direct** financial burden to patient
  - Cost Modeling to predict **indirect** costs of childcare, lost wages, meals, and lodging)

Future AI Solution: Features and Functionality

- Consents
  - Consent Translation
  - Reduction in Language Complexity
  - Interactive trial education bot
  - Mixed media content creation
- Minimizing Deviations
  - On-study patient navigator chat bot for patients and clinical research coordinator.

Deep Dive 2A: Precision Cohorts: AI Tools for Effective Clinical Trial Enrollment

Michael Farkouh, MD, MSc

Associate Dean, Research and Clinical Trials

Professor, Cardiology

Cedars-Sinai Health System

Los Angeles, CA

Cedars Sinai

Disclosures

Research grant support from Novo Nordisk, Astra Zeneca

Artificial intelligence for optimizing recruitment and retention in clinical trials: a scoping review

A scoping review analyzed the use of AI in clinical trial recruitment (2004–2023). Findings from 51 studies highlighted AI's potential to enhance efficiency, reduce costs, improve accuracy, and boost patient satisfaction with Oncology emerging as the most studied area

Challenges included issues with data privacy, transparency, bias, and sample quality

Future research should focus on standardized outcomes and improved methodologies to validate AI's effectiveness in this domain

Benefits of AI tools

- Increased efficiency, cost savings, and improving recruitment
- Improving accuracy and predictive power
- Create user-friendly interfaces and improve patient satisfaction
- Allow remote access and monitoring

Drawbacks of AI tools

- Unreliability of results under data constraints
- Limited generalizability
- The complexity of algorithms and blind trust issues
- Insufficient external validation data and lack of universal guidelines
- Adds additional costs

Ethical Risks

- Privacy and data security
- Informed consent
- How do patients control their data

Cedars- Sinai Deep 6 AI makes Clinical Trials Acceleration Software

We use AI on medical records, to find more, better-matching patients for clinical trials in minutes, not months

Expand care options

Patients

Offer clinical data/services

Partners

Optimize clinical trials performance and profitability

HCO's

DEEP 6 AI

connecting clinical trial stakeholders on an AI-driven, real-time, real-world intelligence platform

Sponsors CROs

Extend time under patent

Reduce time to market

Beat IBM, HP autonomy in USG contest, leading to $2m in contracts

DisruptorDaily

Top 100 most disruptive companies in the world

SXSW

Win at SXSW 2017 accelerator: enterprise + smart data

FAST COMPANY

Finalist, World Changing Ideas Awards 2020

Forbes AI 50

America's Most Promising Artificial Intelligence Companies

Cedars Sinai

Query the deepest source of EMR data

92% **Of I/E criteria benefit** from unstructured data

20% **Of eligible patients** are found through structured data alone

99% **Of genomics data** are found in clinician notes and reports

<20% Structured EMR Data

- **Demographics:** Sex, Age, Ethnicity, ETC.
- **Diagnosis and Billing Codes:** ICD-CP ICD-PM, CPT, HPCS, LOINC
- **Numerical Values:** Height, weight, BP, ETC.
- **Categorical Values:** Blood Type, Disease stages of diagnosis, ETC.
- **Medications**
- **DE&I**

>80% Unstructured EMR Data

- **Symptoms**
- **Clinical Provider Notes**
- **Encounters**
- **Genomics Reports**
- **Imaging Reports**
- **Lab Reports**
- **Lines of Therapy**
- **Pathology Reports**
- **Post Op Notes**
- **Social History**
- **Wearable Data**
- **Disease specific scales**

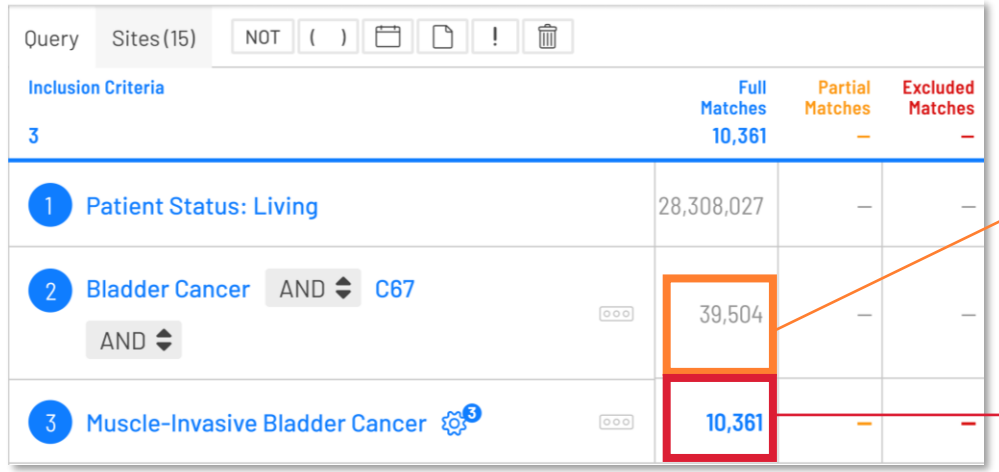
Cedars Sinai

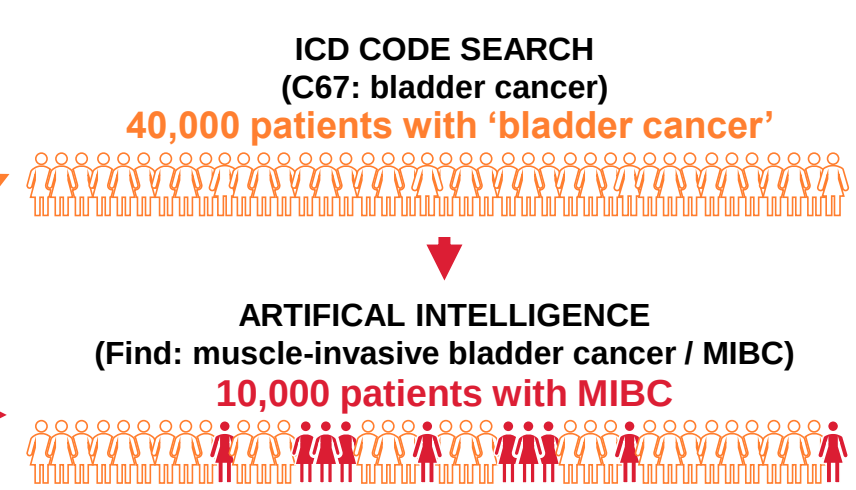
AI Improves Patient Matching Precision

Objective: find patients with ‘muscle-invasive bladder cancer’ (MIBC)

Challenge: code for bladder cancer is C67, no code for MIBC, could appear in notes as:

MUSCLE INVASIVE BLADDER CANCER, MIBC, ADVANCE BLADDER CANCER, INVASIVE BLADDER CANCER; HIGH GRADE BLADDER CANCER, ADVANCED UROTHELIAL BLADDER CANCER, ETC.





*Screenshots are of Deep 6 AI software with MIMIC data (no PHI is shown)

AI pinpoints MIBC 4x more precisely than searching EMR coded data

Cedars Sinai

AI Improves Matching & Validation Speed

Manual chart review

4

hours/patient

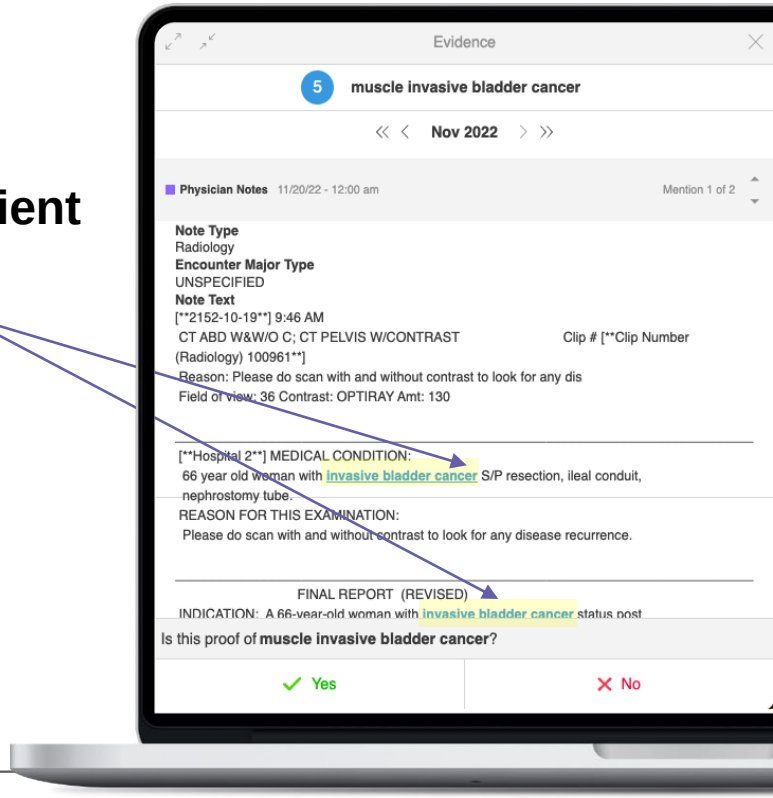
1 MIBC patient

AI assisted review = Validation is 100x faster

<2

minutes/patient

SITE PORTAL



*Screenshots are of Deep 6 AI software with MIMIC data (no PHI is shown)

Cedars Sinai

Next-generation Patient Matching Solutions

STUDY OPTIMIZATION

Protocol Assessment

Assess cohort availability & DEI based on each protocol's inclusion and exclusion criteria

Cohort optimization

Enhance matching precision by working with our clinical and AI experts to optimize the matching query for every cohort

SITE NAVIGATION

Site feasibility

Know whether or not your health system has an adequate patient population for your study and where those patients are located

Deep 6 AI Research Ecosystem

Participate in clinical trials brought by Deep 6 AI pharma partners

Site support

Improve site adoption by having our team manage onboarding, training, and ongoing site support

RECRUITMENT ACCELERATION

Patient matching

Accelerate accruals by giving site staff AI-matched, screen-ready patients to validate

Trial recommender

Expand recruitment to treating physicians and enable them to refer patients to in-network study PIs

RWE GENERATION

Chart review

Curate custom data sets using AI to precisely identify patient charts

Prospective monitoring & registries

Continuously collect endpoint data on a precise patient cohort for research studies

Next-gen RWE studies

Access the entire EMR for deeper clinical data to conduct more precise RWE studies

Benefits of Deep 6

- Can help identify the approximate number of patients seen within a certain time frame that meet a certain criteria.
- Valuable for planning grants and assessing clinical trial feasibility
- Enables pre-screening for study eligibility by reviewing charts based on predefined criteria (with IRB approval)
- Effective for narrowing patient pools
- Has the potential to be helpful to integrate with other centers and work collaboratively

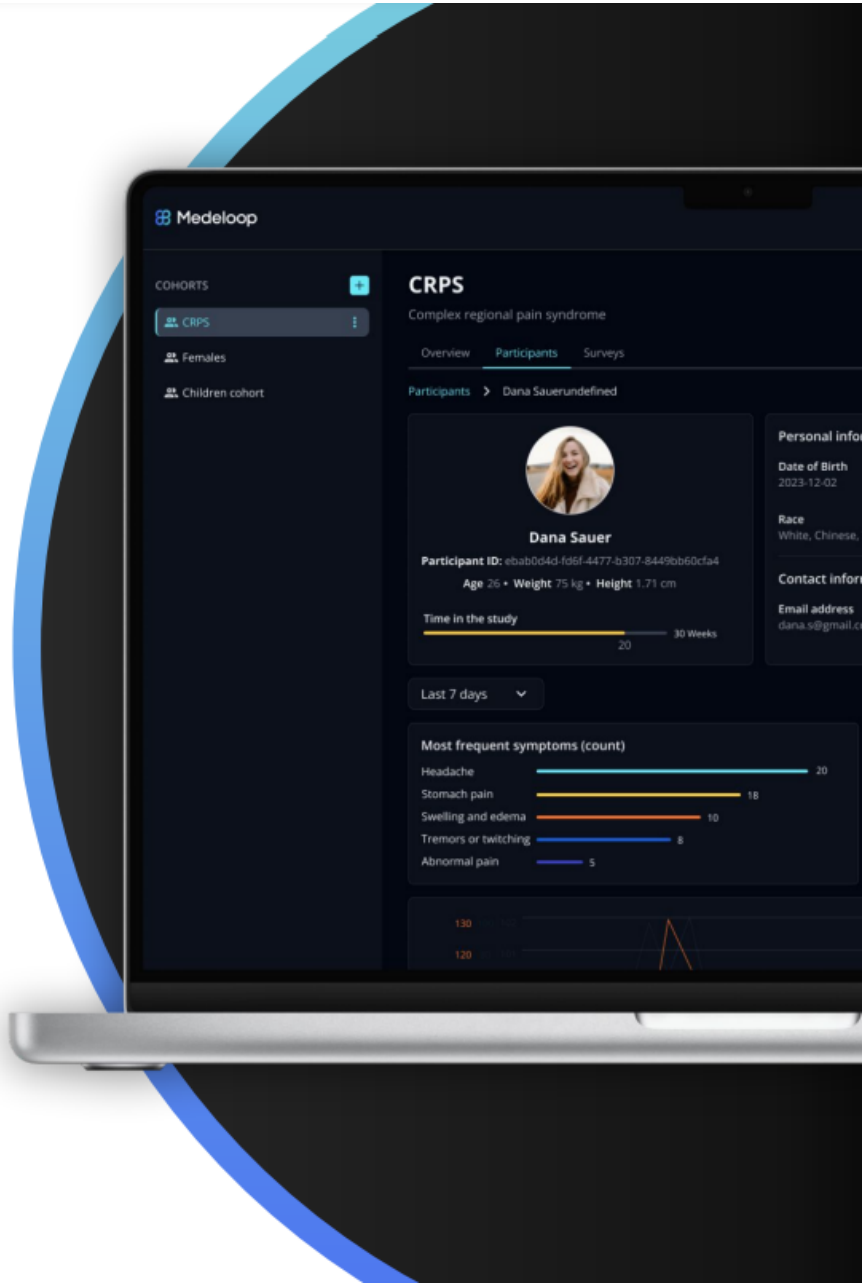
Drawbacks of Deep 6

- Can overcall a condition
- Not as helpful with highly prevalent conditions
- Data exchange challenges include timeline conflicts and non-clinical interpretations, which may misclassify information
- Departments within the hospital that rely on standardized text and non-readable PDFs pose additional challenges for data extraction

Medeloop

Accelerating Medical Research with an End-to-End, AI-Driven Platform

www.medeloop.ai

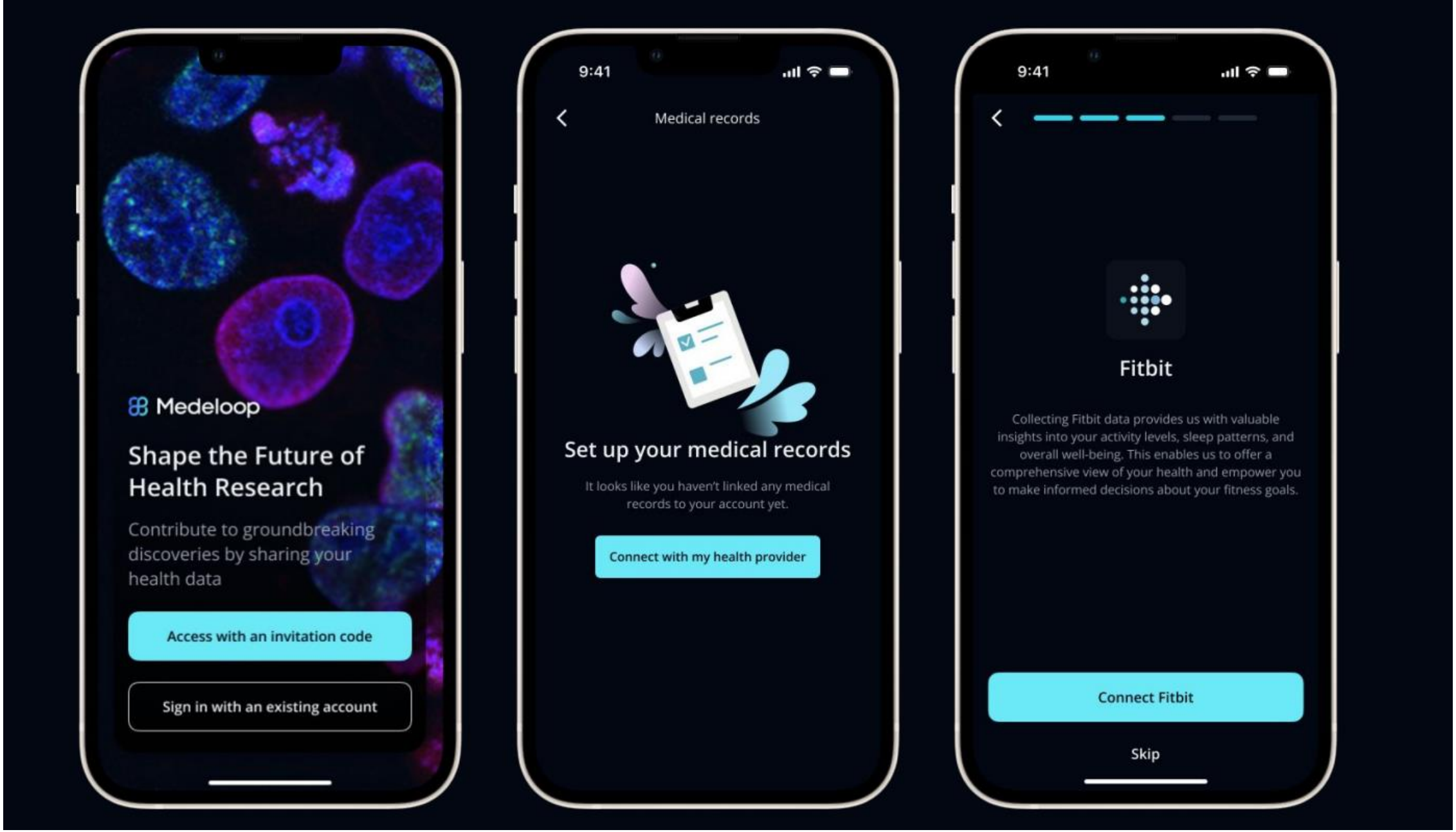


Cedars Sinai

THE SOLUTION

End-to-End Solution for Healthcare Research, QI and Trials

Participant Mobile App


Medeloop
Shape the Future of Health Research
Contribute to groundbreaking discoveries by sharing your health data
Access with an invitation code
Sign in with an existing account
9:41
Medical records
Set up your medical records
It looks like you haven't linked any medical records to your account yet.
Connect with my health provider
9:41
Fitbit
Collecting Fitbit data provides us with valuable insights into your activity levels, sleep patterns, and overall well-being. This enables us to offer a comprehensive view of your health and empower you to make informed decisions about your fitness goals.
Connect Fitbit
Skip