ISCTM
International Society for CNS Clinical Trials and Methodology

Automated Monitoring of Dosing & Other Sessions in Psychedelic Trials Using LLMs

Atul R. Mahableshwarkar MD, DLFAPA

Helus Pharma

Disclosures

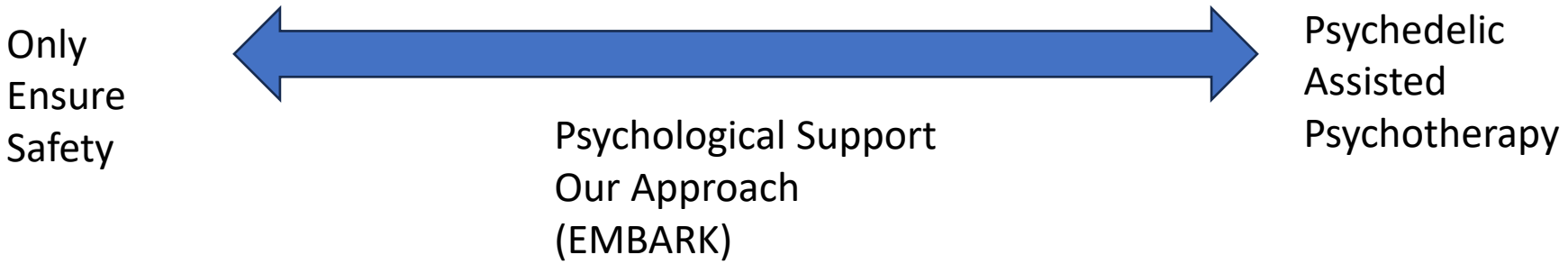
- Full-time employee of Helus Pharma
- Own stock in J&J, Emalex and Evozyne

- Oversight of Session Monitors conducted in collaboration with mpathic - lead by VP of Clinical AI, Megan Greenlaw, MA

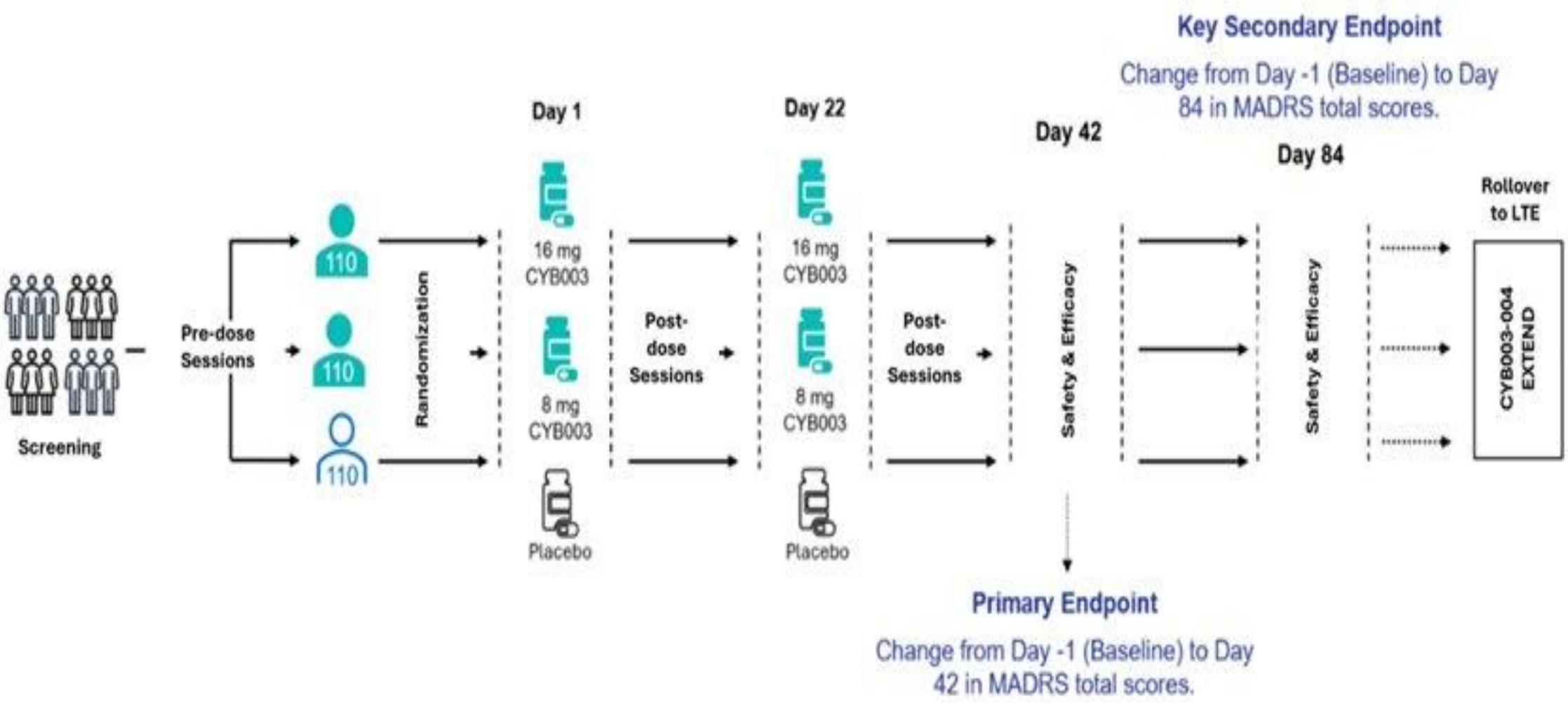
Why Monitor Sessions

- Ensure participant safety during dosing sessions
  - SIB
  - Session monitor misconduct
- Monitor for AEs
- Ongoing assessment of adherence to psychological support model
  - Across sites/session monitors
  - During all sessions by individual session monitors

Range of Psychological Support



Study Design



Some Considerations

- How many preparatory and post-dose dose support sessions
  - ISCTM psychedelics working group consensus
    - Minimum 1 session pre-dose
    - Minimum 1 session post-dose
- How to ensure consistency of support
- Session room layout
  - Furniture
  - Decoration/art etc.
- Participant comfort

What is EMBARK

- A transdiagnostic approach to provide psychological support during dosing with psychedelics
  - Can be tailored for use across diagnoses and/or molecules
- Utilizes existing, accepted practices, while also giving psychedelic-specific training and recommendations for intervention
- EMBARK models six potential domains representing the spectrum of potentially beneficial psychedelic treatment experiences

Approach was deployed in Phase 2 trial and data utilized to develop adherence standards for Phase 3 trials

Six Clinical Domains

Existential-Spiritual
Translating psychedelic-occasioned spiritual experiences into resources for healing.

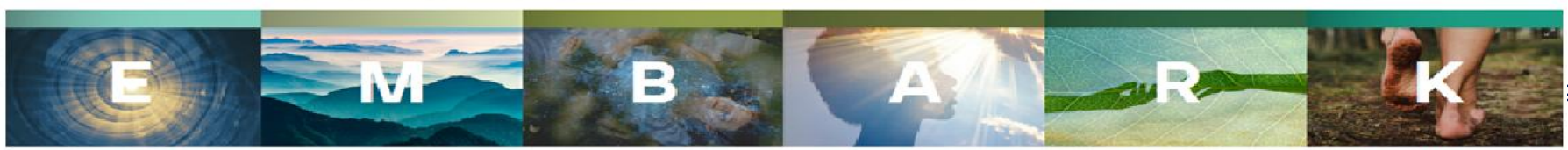
Mindfulness
The practice of focusing awareness on inner experience

Keeping Momentum
The process of integration and translating altered states into altered traits

Body-Aware
Knowing the body can act as a doorway into deep healing

Relational
A unique capacity to dissolve boundaries between self and other

Affective-Cognitive
Supports participants with identifying, welcoming, and experiencing affective (emotional) states more fully



- Session monitors are expected to have a calm, unhurried, supportive presence while maintaining boundaries

AND NOT BE

- Rigid, judgmental, zoned out and artificial

Why Use AI to Monitor Sessions

The challenge

- High-risk trials generate hundreds of hours of session audio/video
- Manual review alone is resource-intensive, inconsistent, and difficult to scale
- Safety and fidelity risks may emerge between scheduled reviews

What AI enables

- **Scalable detection** of clinically relevant behaviors across *all* sessions
- **Consistent application** of adherence criteria across sites and facilitators
- **Early identification** of potential safety, boundary, or fidelity concerns
- **Targeted human review**, rather than exhaustive manual screening

AI supports clinician-led oversight by narrowing attention to moments that matter.

Approach to AI Safety Monitoring

Where this AI support system has been used

- High-risk, highly regulated environments where humans must review hours of audio/video, including:
    - Psychedelic and CNS clinical trials
    - Medical visits and surgical care settings (the OR)
    - Psychotherapy and behavioral health care
    - Other clinical and institutional contexts requiring strong privacy and governance

Relevance for psychedelic trials

- Careful monitoring of facilitator behavior during all sessions
- Consistent application of psychological support models
- Cross-check to ensure accurate and complete reporting of potential abuse-related AEs
- Scalable review without reducing clinical oversight

Designed to support safety-focused review in settings where accuracy and trust are critical

From Clinical Fidelity to Interpretable Adherence Flags

Clinical grounding

- Adherence criteria were derived from the EMBARK psychological support model
- Clinicians translated EMBARK fidelity standards into observable, reviewable behaviors
- Behavioral definitions were reviewed and aligned with expert clinical judgment to ensure real-world validity
- AI models were trained on the expert clinical judgement
  - Benchmarked until they met ≥ 80% IRR with expert clinicians
  - Ongoing checks of AI model performance to ensure ≥ 80% IRR

Examples of non-adherence flags (illustrative)

- Boundary violation
- Missed emotional support
- Process deviation (Language inconsistent with EMBARK guidance)

What reviewers see

- **Timestamped flag** linked to transcript and audio/video
- **Behavioral description** tied to fidelity criteria (not diagnostic labels)
- **Clinician-in-the-loop review** to confirm non-adherence and determine next steps

How Sessions are Reviewed

Each recorded session is transcribed and reviewed utilizing 200+ proprietary AI models

- Identifying adherence to the psychological support model
- Clinically defined cues are used to assess adherence to the EMBARK model
- To identify misconduct and clinical risk/AEs

The system flags sessions with specific time stamps – connected to both the transcript and audio or video file that is reviewed by a clinician

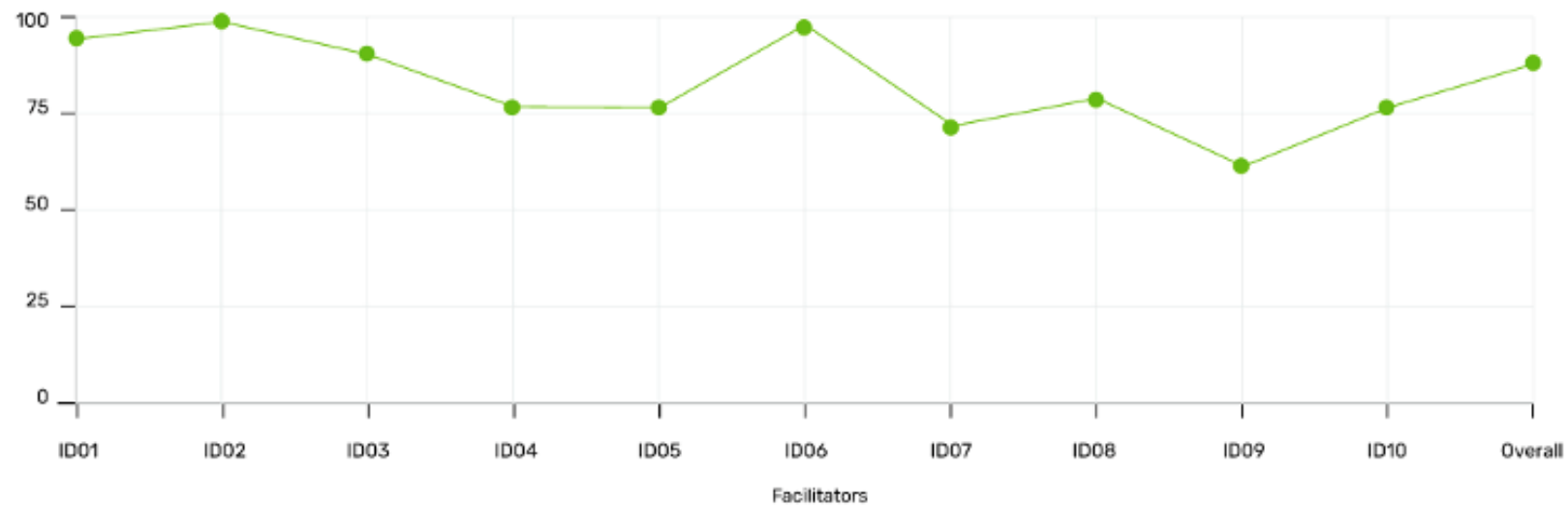
- If risk is detected, trial contact is notified within set hours for safety action (24, 48, 72 hrs, etc. – depending on trial)

Over time, the AI system will demonstrate where more specificity is needed and/or when certain constructs evolve over time

- Dashboard
  - Protocol Adherence
    - By Trial
    - By Site
    - By Facilitator
  - Safety Detections
- Reports

Protocol Adherence by Facilitator

Overall Average Adherence Score by Facilitator



Facilitator	Preparation		Administration		Integration		Overall Average Score
	Score (%)	n	Score (%)	n	Score (%)	n	
ID01	100	4	100	8	75	2	**90%**
ID02	75	1	100	1	100	1	**98%**
ID03	100	5	75	3	100	1	**88%**
ID04	75	1	100	2	0	1	**78%**

Adherence Review Summary: CYB-003 Trial

The CYB-003 trial involved a structured therapeutic protocol, with support provided pre-dose, during dosing and post-dose. To assess fidelity to this protocol, adherence was evaluated at two points:

- **Phase 1:** After the first three preparation sessions and initial dosing session
- **Phase 2:** After receipt of all available session recordings

To date, **442 session recordings** have been reviewed via the mpathic platform (note: this number may include some duplicates).

Adherence Findings

- Cases involving potential adherence issues identified by the AI model were reviewed by human experts
- After expert review as necessary session monitors are called to discuss potential adherence issues and suggestions for improvements for future sessions

Important Details on AI Support

- Who/What is hallucinating
  - The participant after receiving IP?
  - The AI system?
- Since the system operates on transcript-anchored analysis
  - Which is verified by expert human clinician human verification
  - Hallucinated content does not influence review or clinical decisions
  - A percentage of transcripts are reviewed regardless of absence of flags
- **The end goal is to support clinician-led review—augmented by AI—of fidelity, clinical risk, and potential misconduct, while preserving participant safety and trial rigor**

Acknowledgments

Bill Brennan

Mellissa Field

Amir Inamdar

Alex Kelman

Megan Greenlaw

Alison Cerezo

Danielle Schlosser