From Monitoring to Precision: Machine-Learning Approaches to Personalizing Predictors of Psychotherapy Process & Outcome

Mor Keinan-Bar* Elad Refoua* Yonatan Bilu Dan Sayda Yogev Kivity Eshkol Rafaeli Dana Atzil-Slonim

Dept. of Psychology, Bar-Ilan University (ARLAB) • KI Research Institute, Hod HaSharon • School of Psychological Sciences, Tel Aviv University | * equal contribution

SPR 57th Annual International Meeting

Osaka · 2026

Submission #236232

Work in progress · descriptive overview · no outcome results yet

◆ THE QUESTION WE SET OUT TO ANSWER ◆

Can routine outcome monitoring — questionnaires only, no session text — be turned into a personalized forecast of who improves, in which domain, and who is at risk of deterioration?

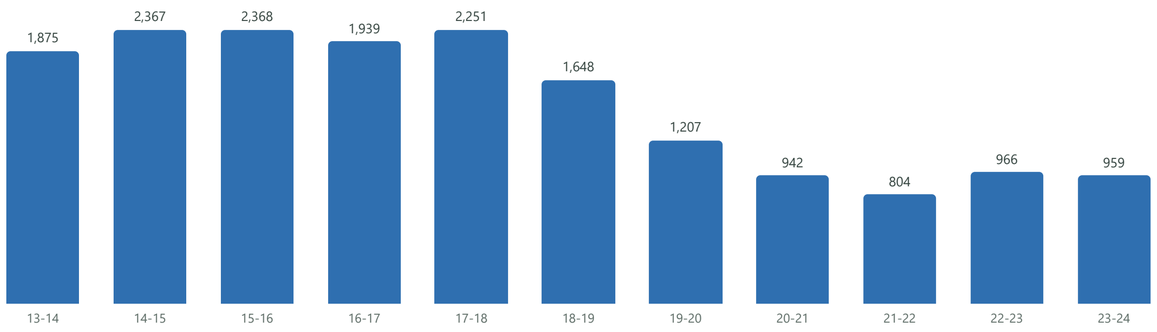
We set out to answer this with **11 years** of the clinic's routine data — a broad, multi-informant, session-by-session battery.

OUR APPROACH a two-phase machine-learning program — **benchmark** a per-client outcome forecast at **session 3-5**, then **personalize** it by domain & flag deterioration

17,326	924	693	504	129,894	22	3	11
sessions	dyads	clients	therapists	questionnaire administrations	measures	informants	years 2013-24

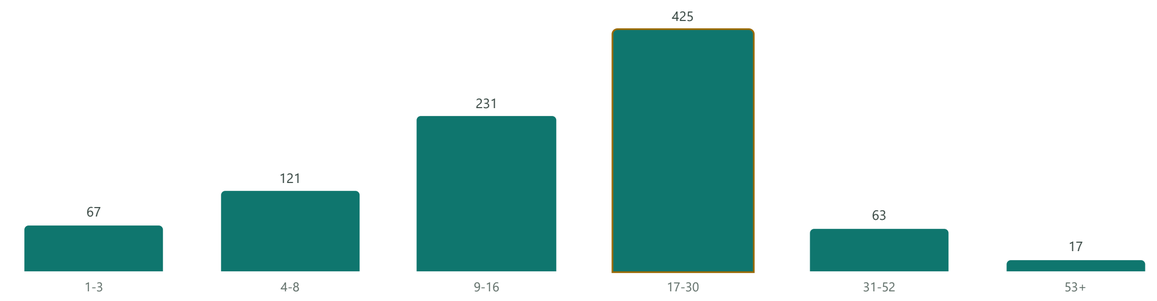
Sessions per year

17,326 across 11 years (COVID/war dip visible)



Treatment length

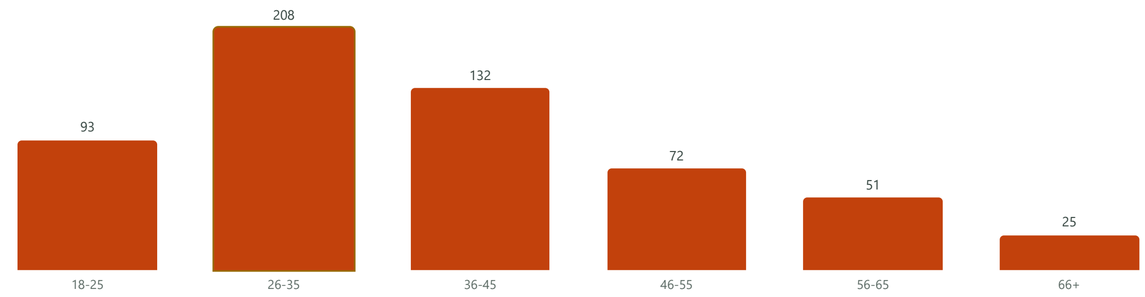
sessions per dyad · median 18 (IQR 10-24)



most dyads run **17-30** sessions

Client age distribution

consented clients · broad adult range



peak **26-35** · spanning 18 to 79

Three informants

same constructs, different eyes (dyads with data)

enables agreement & empathic-accuracy features

Who the clients are

consented analytic sample

37.9	59.6%	18-79
mean age (SD 13.2)	female	age range

a broad adult outpatient population

Returning clients

clients seen across multiple academic years

10	412	1.5
max years, one client	dyads with pre & post OQ-45	mean clients per therapist/yr

The measurement battery — 22 instruments, full names

informant (dot): client / therapist / therapist-on-client · items · consented N | right: how the battery feeds the model

SESSION

before / after every session

ORS	Outcome Rating Scale	4 it	699
HSCL	Hopkins Symptom Checklist	11 it	699
WAI-6	Working Alliance Inventory	6 it	622
POMS	Profile of Mood States	12 it	621
SES	Session Evaluation Scale	1 it	409
Rupture	Rupture self-report item	1 it	622

The longitudinal core — repeated before/after every session, giving each client a dense within-treatment trajectory.

BASELINE

once per academic year

OQ-45	Outcome Questionnaire-45	45 it	599
BDI	Beck Depression Inventory	21 it	598
IIP	Inventory of Interpersonal Problems	32 it	593
DERS-18	Difficulties in Emotion Regulation Scale	18 it	560
ERQ	Emotion Regulation Questionnaire	10 it	497
PID-5	Personality Inventory for DSM-5	25 it	347
SPIN	Social Phobia Inventory	17 it	448
SE	Rosenberg Self-Esteem Scale	10 it	449
SWLS	Satisfaction With Life Scale	5 it	326
SHEEHAN	Sheehan Disability Scale	3 it	189
IPDE	Int'l Personality Disorder Examination	77 it	143

THERAPIST

self-report

WAI-6 (T)	Working Alliance Inventory (therapist)	6 it	600
IIP (T)	Inventory of Interpersonal Problems (therapist)	32 it	374
DERS (T)	Difficulties in Emotion Regulation (therapist)	18 it	328

The therapist's own interpersonal style & emotion regulation — the basis for therapist-effect predictors.

CROSS-RATER

therapist-on-client

tc-ORS	Outcome Rating Scale (therapist-rated)	4 it	522
tc-POMS	Profile of Mood States (therapist-rated)	12 it	482

The therapist's view of the same client — enables client-therapist **agreement** & **empathic-accuracy** features, among the strongest known predictors.

FROM BATTERY TO FORECAST

22 instruments
3 informants · 11 years
129,894 questionnaire administrations

↓

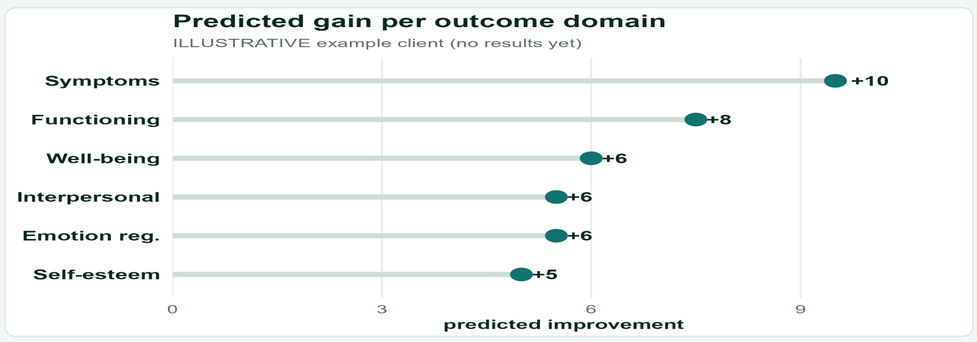
Tree-ensemble model
missing-tolerant · leave-one-year-out CV · honest emulation

↓

6 personal forecasts
per client, by outcome domain — plus a deterioration flag

HOW WE WILL ANSWER IT

a two-phase machine-learning program · questionnaires only · figures are illustrative (no results yet)



Phase 1 · Benchmark

At session **3-5**, baseline + early sessions forecast each client's outcome, and we establish how accurately it can be done.

Phase 2 · Personalize

Trajectory-predicts-trajectory, data-driven profiles, SHAP explanations, deterioration flags, and per-client predicted gains via interactions.

Forecast domains

Symptoms · Functioning · Emotion reg. · Interpersonal · Self-esteem · Well-being

Strongest predictors

- Early-treatment trajectory (strongest)
- Client-therapist alliance agreement
- Therapist under-estimating alliance
- Rupture recognition → repair
- Therapist empathic accuracy

INSTRUMENT SOURCES **ORS** Miller et al., 2003 · **HSCL** Derogatis et al., 1974 · **WAI** Horvath & Greenberg, 1989 · **POMS** McNair et al., 1971 · **SES** single clinic item · **Rupture** single clinic item · **OQ-45** Lambert et al., 1996 · **BDI** Beck et al., 1996 · **IIP** Horowitz et al., 1988 · **DERS-18** Victor & Klonsky, 2016 · **ERQ** Gross & John, 2003 · **PID-5** Krueger et al., 2012 · **SPIN** Connor et al., 2000 · **SE** Rosenberg, 1965 · **SWLS** Diener et al., 1985 · **SHEEHAN** Sheehan, 1983 · **IPDE** Loranger et al., 1994 · **DERS (T)** Gratz & Roemer, 2004

Work in progress
no outcome results yet

Routine monitoring, re-imagined as a personal forecast.

Contact: eladrefoua@gmail.com · Bar-Ilan University & KI Research Institute, Hod HaSharon