

◆ 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  
sessions

924  
dyads

693  
clients

504  
therapists

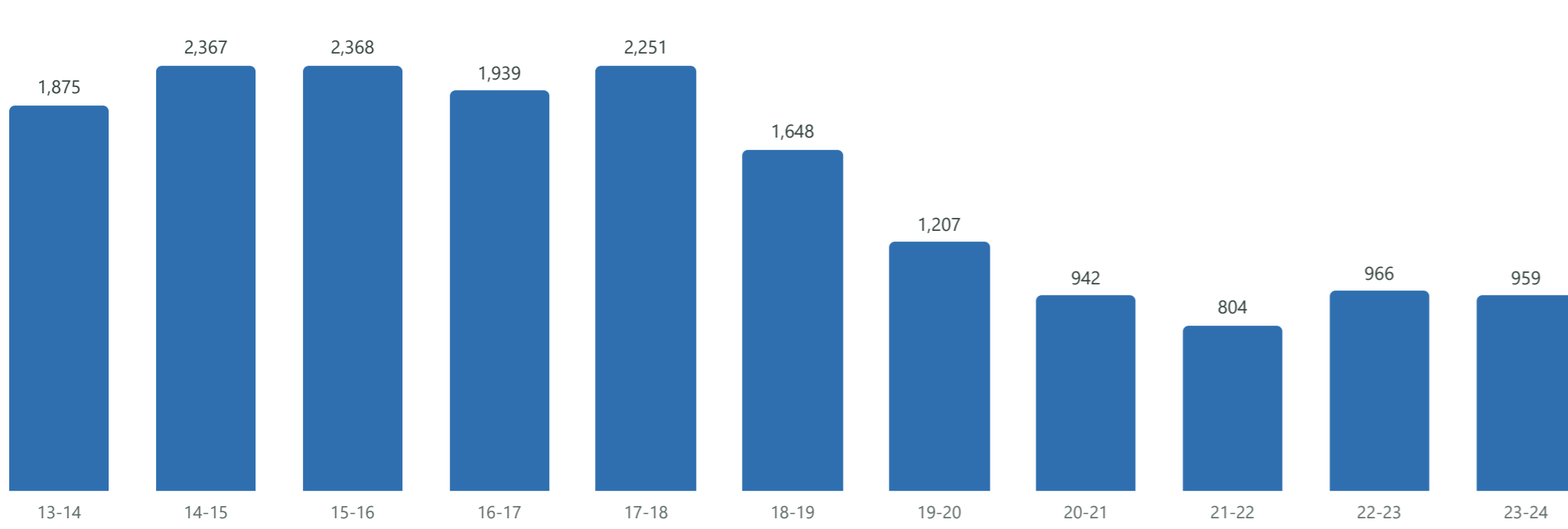
129,894  
questionnaire  
administrations

22  
measures

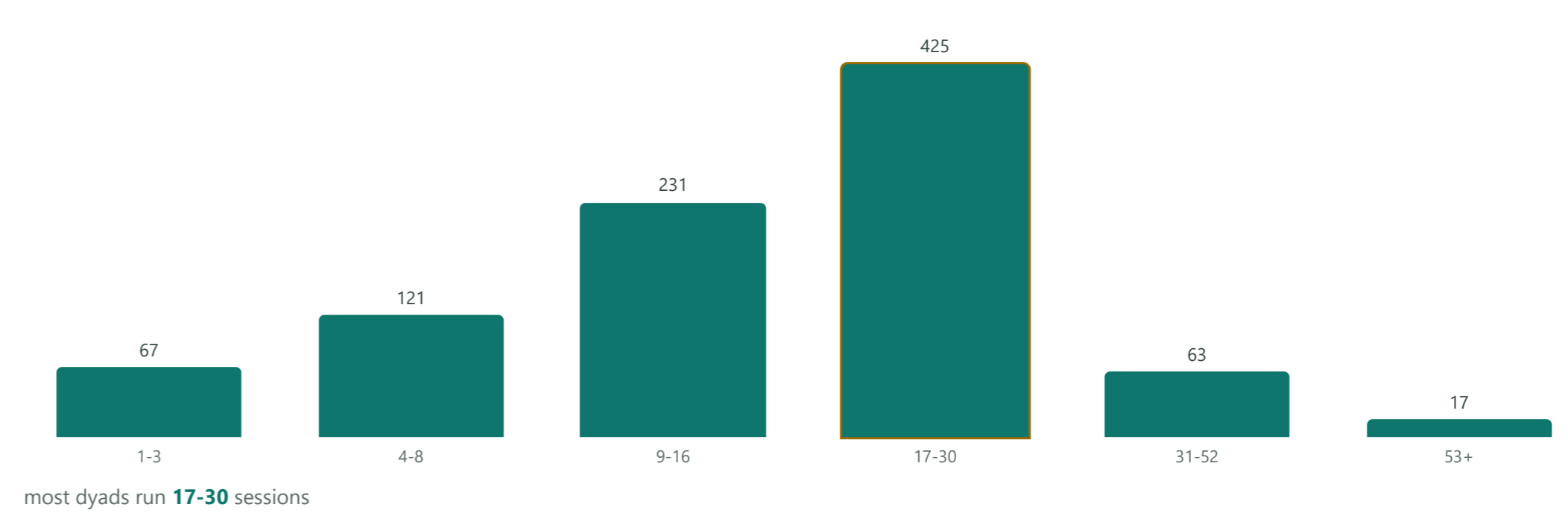
3  
informants

11  
years 2013-24

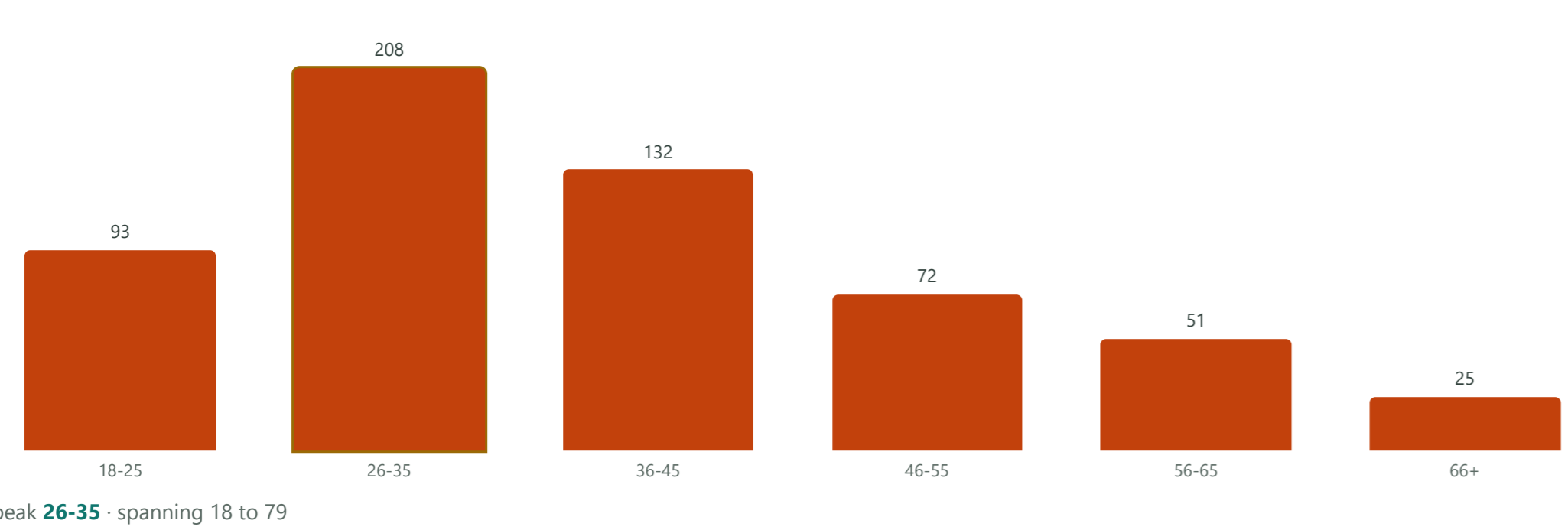
### Sessions per year



### Treatment length



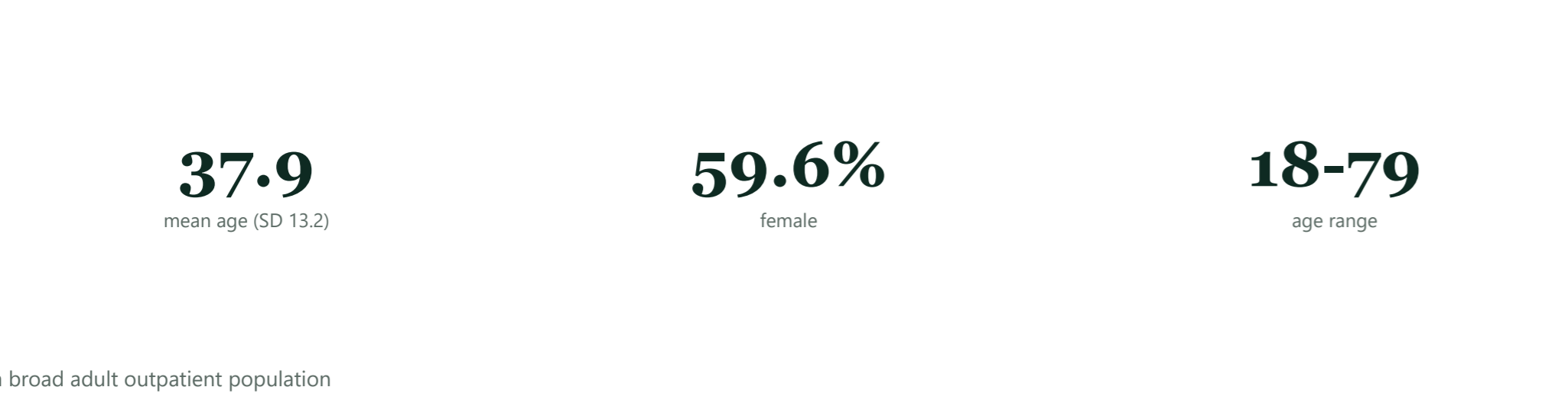
### Client age distribution



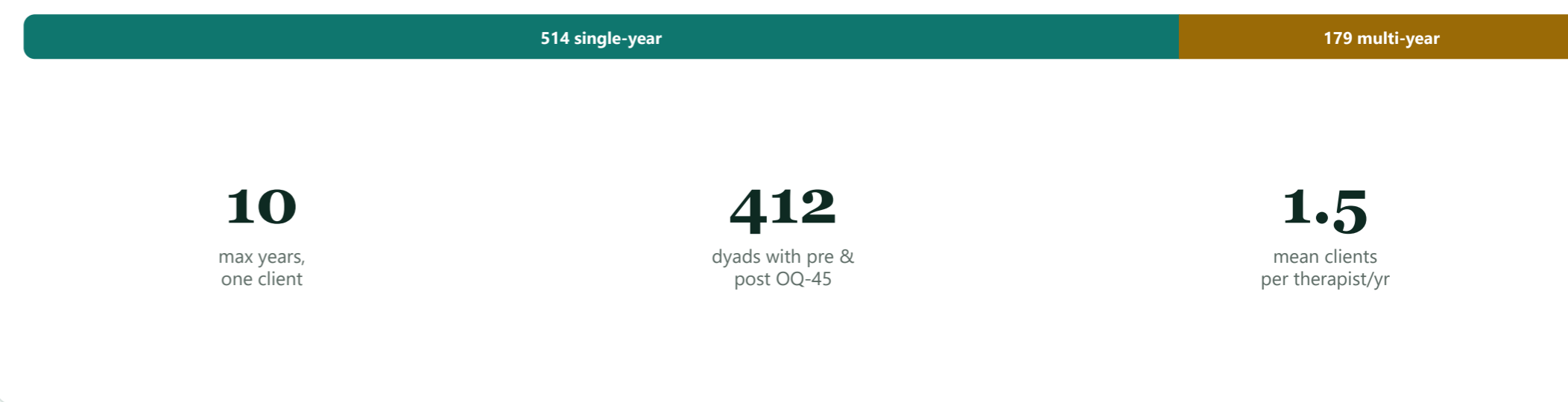
### Three informants



### Who the clients are



### Returning clients

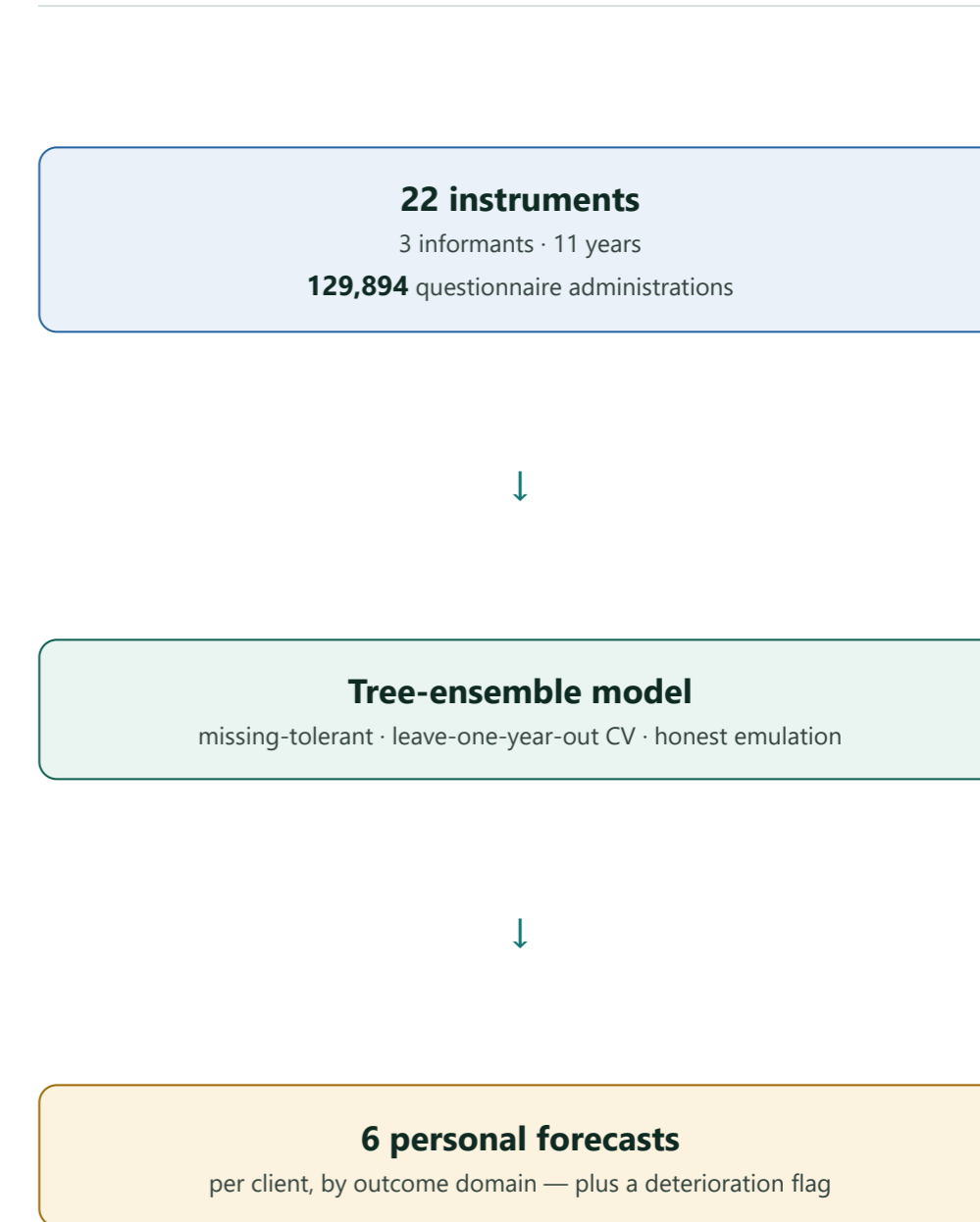


### The measurement battery — 22 instruments, full names

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

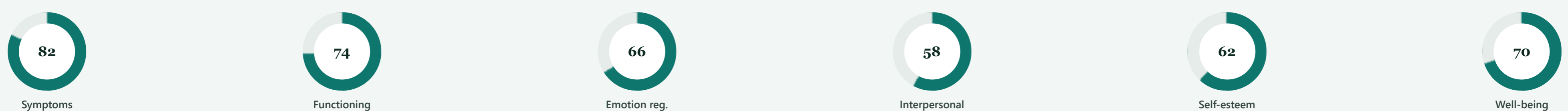
SESSION	BASELINE	THERAPIST
before / after every session	once per academic year	self-report
<ul style="list-style-type: none"> <li>● <b>ORS</b> Outcome Rating Scale (4 it, 699)</li> <li>● <b>HSCL</b> Hopkins Symptom Checklist (11 it, 699)</li> <li>● <b>WAI-6</b> Working Alliance Inventory (6 it, 622)</li> <li>● <b>POMS</b> Profile of Mood States (12 it, 621)</li> <li>● <b>SES</b> Session Evaluation Scale (1 it, 409)</li> <li>● <b>Rupture</b> Rupture self-report item (1 it, 622)</li> </ul>	<ul style="list-style-type: none"> <li>● <b>OQ-45</b> Outcome Questionnaire-45 (45 it, 599)</li> <li>● <b>BDI</b> Beck Depression Inventory (21 it, 598)</li> <li>● <b>IIP</b> Inventory of Interpersonal Problems (32 it, 593)</li> <li>● <b>DERS-18</b> Difficulties in Emotion Regulation Scale (18 it, 560)</li> <li>● <b>ERQ</b> Emotion Regulation Questionnaire (10 it, 497)</li> <li>● <b>PID-5</b> Personality Inventory for DSM-5 (25 it, 347)</li> <li>● <b>SPIN</b> Social Phobia Inventory (17 it, 448)</li> <li>● <b>SE</b> Rosenberg Self-Esteem Scale (10 it, 449)</li> <li>● <b>SWLS</b> Satisfaction With Life Scale (5 it, 326)</li> <li>● <b>SHEEHAN</b> Sheehan Disability Scale (3 it, 189)</li> <li>● <b>IPDE</b> Int'l Personality Disorder Examination (77 it, 143)</li> </ul>	<ul style="list-style-type: none"> <li>● <b>WAI-6 (T)</b> Working Alliance Inventory (therapist) (6 it, 600)</li> <li>● <b>IIP (T)</b> Inventory of Interpersonal Problems (therapist) (32 it, 374)</li> <li>● <b>DERS (T)</b> Difficulties in Emotion Regulation (therapist) (18 it, 328)</li> </ul>
<p>The longitudinal core — repeated before/after every session, giving each client a dense within-treatment trajectory.</p>	<p>The therapist's own interpersonal style &amp; emotion regulation — the basis for therapist-effect predictors.</p>	<p>The therapist's view of the same client — enables client-therapist agreement &amp; empathic-accuracy features, among the strongest known predictors.</p>

### FROM BATTERY TO FORECAST



### 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.

#### Method

- Curated, not blind literature-grounded predictors
- Prediction, not inference out-of-sample accuracy is the target
- Missing-tolerant trees gradient-boosted ensembles
- Leave-one-year-out CV COVID / war years stress-test it
- Temporally honest only data available at the prediction point

