PSILOGOS

Research Article

open-access

Individualized Session Allocation Under Heterogeneous Psychotherapy Response: A Monte Carlo Study of Early Clinical Change

Valeria Frighi

ABSTRACT

Fixed psychotherapy courses are commonly specified before the clinician has observed whether a person is improving rapidly, gradually, or only after an initial delay. The clinical consequences of using early symptom change to select treatment length remain difficult to evaluate in ordinary trials because each participant receives only one duration. This study examined whether a four-session allocation rule could preserve composite recovery while reducing completed sessions relative to fixed 8-, 16-, and 24-session courses. A prespecified Monte Carlo model generated 12,000 synthetic adult outpatient trajectories over 24 sessions. Initial symptom burden, functional impairment, chronicity, clinical complexity, and one of four latent response patterns determined repeated symptom and functioning scores. Logistic models, developed in 6,000 trajectories, estimated the probability of composite recovery by sessions 8 and 16 from initial characteristics and proportional symptom change at session 4. The remaining 6,000 trajectories received an 8-session course when the estimated probability of recovery by session 8 was at least 0.70; otherwise they received 16 sessions when the estimated probability of recovery by session 16 was at least 0.55, and 24 sessions in the remaining cases. Composite recovery required a symptom score of 30 or lower, a symptom reduction of at least 50%, and an 8-point functional gain. **Results:** The allocation rule achieved 86.0% composite recovery (95% bootstrap interval 85.1–86.8) with a mean of 13.6 sessions (13.5–13.7). Fixed 16 sessions achieved 86.9% recovery (86.0–87.7) with 16.0 sessions, whereas fixed 8 sessions achieved 41.1% recovery (39.9–42.3). The allocation rule assigned 30.1% of participants to 8 sessions, 69.9% to 16 sessions, and 0.02% to 24 sessions. In this synthetic setting, early symptom change can support a shorter average course with recovery close to that of a fixed 16-session schedule. These findings establish operating characteristics of the rule, not clinical effectiveness. Prospective trials should test whether the observed balance between recovery and session use holds in routine care.

Keywords: psychotherapy duration; early response; individualized care; treatment allocation; Monte Carlo study; mental health services

Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, United Kingdom
valeria.frighi@psych.ox.ac.uk

– **Published:** 20 November 2024

1. INTRODUCTION

Mental health disorders contribute substantially to disability, reduced participation in work and education, and health-service expenditure.[1,2] Psychotherapy is effective for many common disorders, yet decisions about treatment length are often made with limited knowledge of the time course of a particular person's response. A fixed number of appointments is administratively simple, but it places rapid responders and people with delayed improvement on the same calendar. This is important in services where each additional appointment for one patient may postpone access for another.

The relation between completed sessions and outcome has been discussed for decades. Early dose-effect accounts described a negatively accelerating relation, in which a considerable proportion of improvement appeared early in care.[3] Later quantitative work found that more sessions were associated with greater improvement in depression, while also emphasizing that the average association did not determine the duration appropriate for an individual patient.[4] A systematic review of routinely delivered psychological therapies reached a similar conclusion: the shape of the association differed across settings, diagnoses, and definitions of improvement.[5] The good-enough-level literature further suggests that some people leave treatment after achieving a personally sufficient benefit, whereas others require continued work before comparable improvement is evident.[6]

Direct duration comparisons remain sparse and heterogeneous. Trials have compared brief and standard cognitive behavioural therapy for panic disorder,[7–9] intensive and weekly cognitive therapy for post-traumatic stress disorder,[10,11] brief and full internet-delivered cognitive behavioural therapy for depression,[12,13] and short versus long psychodynamic treatment for mood and anxiety disorders.[14–18] Other studies have contrasted session frequency,[19] duration of dialectical behaviour therapy,[20] and prolonged-exposure formats.[21,22] Internet-delivered treatment has also been tested with a restricted cognitive component,[25] and duration selection has been examined in complex personality presentations.[26,27] The estimates are difficult to synthesize because the studies vary in diagnosis, treatment modality, delivery format, treatment intensity, endpoint selection, and follow-up window. Systematic-review reporting and certainty assessment make these differences explicit.[28,29] One review of randomized duration comparisons found 19 trials involving 3,447 participants and concluded that the certainty of the evidence was very low.[30] That finding identifies an important clinical problem but does not show how a service should use information emerging during treatment.

The present research question is therefore different: *can a prespecified decision rule that uses symptom change after four sessions retain recovery close to a fixed 16-session course while decreasing the average number of completed sessions?* The question concerns allocation after treatment has begun, rather than a universal contrast between short and long psychotherapy. It is well suited to a synthetic cohort because every generated participant can be evaluated under several possible course lengths without implying that an individual person actually received more than one treatment duration. The study does not estimate effectiveness in patients; it evaluates the internal behaviour of a transparent decision rule under stated assumptions.

Early change is a plausible allocation signal. Patient profiling and risk stratification have shown that treatment response is associated with initial severity, complexity, and symptom trajectory.[31,32] Within routine outcome monitoring, information obtained during the first appointments may be clinically more useful than an average duration estimate calculated before care begins. At the same time, early change should not be mistaken for a complete prognosis. A slow start may reflect delayed engagement, unstable life circumstances, comorbidity, or a response curve that becomes steeper later in treatment. Any allocation rule should therefore be judged against fixed courses not only for symptom improvement but also for functional change and for the proportion of patients who would be assigned too few sessions.

Simulation offers a disciplined way to study this problem when individual participant records with repeated assessments and several alternative durations are unavailable. A useful simulation makes the generating assumptions explicit, uses a fixed seed, separates model development from model evaluation, and reports quantities that permit the reader to reconstruct the decision process.[33,34] Prediction rules also require more than discrimination alone. Their calibration, intended use, and clinical consequences matter, as emphasized by prediction-model reporting and appraisal guidance.[35–37] The present work therefore specifies a heterogeneous response model, fits allocation models only in a development subset, evaluates the completed course in an independent subset, and reports both recovery and session burden.

The synthetic cohort was deliberately constructed to contain four clinically recognizable temporal patterns: rapid response, gradual response, delayed response, and limited response. It also incorporates variation in initial symptom burden, functioning, chronicity, and clinical complexity. These features permit a direct examination of a concern that fixed-course comparisons cannot resolve: whether a person who has not improved sufficiently at session 4 should always continue to a longer fixed course, or whether their profile supports a different treatment-length decision. The analysis adds a sensitivity exercise in which the prevalence of delayed response and the impact of complexity are changed over plausible ranges. The result is a focused quantitative study of how an early-response rule performs, rather than a claim that one psychotherapy duration is uniformly preferable.

2. METHODS

This was a Monte Carlo study using a synthetic cohort of 12,000 adult outpatient treatment trajectories. The unit of analysis was a participant-course pair. Every participant had potential symptom and functioning values at each completed session from 0 through 24. The main estimand was the difference in the proportion achieving composite recovery between the early-response allocation rule and a fixed 16-session course. The co-primary service quantity was the difference in mean completed sessions. Fixed 8-session and fixed 24-session courses were included to make the recovery–capacity balance interpretable.

No patient records, clinical encounters, or identifiable information were used. The cohort was generated from a prespecified stochastic model and should not be interpreted as representing the prevalence of a disorder, the effect of a named psychotherapy, or the performance of any health system. The simulation design follows recommendations that a medical-statistical simulation should state its data-generating mechanism, estimands, evaluation sample, and uncertainty calculation.[33,34]

2.1. Synthetic cohort

Initial symptom severity was sampled from a truncated normal distribution with mean 51 and standard deviation 9.5 on a 0–100 symptom score, bounded between 30 and 75. Initial functioning was generated on a 0–100 scale with an inverse association with symptom severity and a decrement for clinical complexity. Chronicity was sampled from a gamma distribution with mean approximately 5.2 years. Clinical complexity was a binary indicator generated from a logistic probability that increased with symptom severity and chronicity. The binary indicator denotes the presence of features likely to complicate progress, such as comorbid symptoms, persistent social stressors, or prior treatment difficulty; it does not correspond to a diagnostic category.

Each participant was assigned one latent response pattern before any session values were generated: rapid response (probability 0.29), gradual response (0.37), delayed response (0.22), or limited response (0.12). Rapid response followed a steep early decline that approached an asymptote. Gradual response improved steadily across the course. Delayed response had minimal change through session 4 and an accelerating decline thereafter. Limited response improved modestly and then levelled off. Complexity reduced the achievable symptom decline over time. Independent normal disturbance was added at each session to reflect measurement variation and within-person fluctuation.

Let Y_{it} denote the symptom score for participant i after t sessions, Y_{i0} denote the initial score, z_i denote the complexity indicator, and $g_i(t)$ denote the response-pattern-specific proportion of potential improvement. The symptom process was

$$Y_{it} = \max\left\{1,\ Y_{i0}\left[1 - g_i(t) + 0.060 z_i\left(1 - e^{-0.12t}\right)\right] + \varepsilon_{it}\right\}, \tag{1}$$

where $\varepsilon_{it} \sim N(0, 1.55^2)$. The four functions $g_i(t)$ were respectively $0.98(1 - e^{-0.20t})$, $0.84(1 - e^{-0.115t})$, $0.86[1 - e^{-0.115\max(t-4,0)}]$, and $0.40(1 - e^{-0.075t})$. Functional status was generated from initial functioning, proportional symptom reduction, complexity, and an independent normal disturbance:

$$F_{it} = F_{i0} + 20\left(\frac{Y_{i0} - Y_{it}}{Y_{i0}}\right) - 1.6 z_i + \eta_{it}, \tag{2}$$

where $\eta_{it} \sim N(0, 1.4^2)$. Values were restricted to the scale range. The numerical quantities in these equations define the simulated setting; they are not estimated from clinical observations.

2.2. Composite recovery

The outcome combined symptom and functioning criteria because symptom reduction without a corresponding functional gain may have limited clinical value. A participant was classified as recovered at session t only when all three conditions held:

$$R_{it} = 1 \quad \text{if} \quad Y_{it} \le 30, \quad \frac{Y_{i0} - Y_{it}}{Y_{i0}} \ge 0.50, \quad \text{and} \quad F_{it} - F_{i0} \ge 8. \tag{3}$$

The first two conditions require a low end-of-course symptom score and at least a 50% reduction; the third requires an 8-point functional gain. This composite avoids describing a person with a numerical symptom reduction but unchanged functioning as fully recovered. Symptom and functioning outcomes were also reported separately to show whether a policy achieved its recovery rate by a disproportionate change in only one domain.

2.3. Allocation rule

The first 6,000 trajectories formed the development subset and the remaining 6,000 formed the evaluation subset. The split was assigned before fitting either model. Two logistic regressions were fitted in the development subset: one predicted recovery at session 8 and the other predicted recovery at session 16. Both included initial symptom severity, initial functioning, chronicity, complexity, and proportional symptom improvement at session 4. This small predictor set was chosen because it contains information that could, in principle, be available early in a treatment episode. It also avoids constructing a rule that relies on unmeasurable latent response membership.

For an evaluation participant, $\widehat{p}_8$ and $\widehat{p}_{16}$ were the two estimated recovery probabilities. The rule assigned an 8-session course when $\widehat{p}_8 \ge 0.70$. Participants not meeting that criterion were assigned 16 sessions when $\widehat{p}_{16} \ge 0.55$. The remaining participants were assigned 24 sessions.

The thresholds were selected before simulation to prioritize a high probability before truncating the course at 8 sessions while retaining a modest safety margin for the 16-session decision. The rule is expressed as

$$D_i = \begin{cases} 8, & \widehat{p}_8 \geq 0.70, \\ 16, & \widehat{p}_8 < 0.70 \text{ and } \widehat{p}_{16} \geq 0.55, \\ 24, & \widehat{p}_8 < 0.70 \text{ and } \widehat{p}_{16} < 0.55. \end{cases} \quad (4)$$

The policy did not change the symptom-generating equation. It selected the point at which the trajectory was evaluated. This isolation is intentional: it permits the analysis to identify the consequences of selecting duration, rather than attributing a causal effect to the statistical model itself.

2.4. Comparators and statistical analysis

The evaluation subset was assessed under four policies: fixed 8 sessions, fixed 16 sessions, fixed 24 sessions, and early-response allocation. For each policy, composite recovery, mean symptom score at the final completed session, mean change in functioning, and mean completed sessions were calculated. The area under the receiver-operating-characteristic curve was calculated for the session-8 and session-16 prediction models, but recovery and completed sessions were treated as the decision-relevant quantities. This distinction is important because a model may discriminate well while still failing to support a clinically appropriate allocation.[38,40]

Uncertainty intervals were obtained from 2,000 nonparametric bootstrap resamples of the 6,000-participant evaluation subset. Percentile intervals are reported. The sensitivity exercise varied the delayed-response proportion from 12% to 36% and multiplied the complexity penalty by 0.70, 0.85, 1.00, 1.15, or 1.30. For each of the 25 settings, the recovery difference between early-response allocation and fixed 16 sessions, together with sessions saved, was calculated from the prespecified operating model. The design separates development from evaluation but is not an external validation study; further validation would require independently collected clinical cohorts.[39,40]

3. RESULTS AND DISCUSSION

3.1. Cohort characteristics and prediction

The generated cohort contained 12,000 participants. Mean initial symptom severity was 50.94 (SD 9.35), mean initial functioning was 37.17 (SD 8.46), and mean chronicity was 5.23 years (SD 3.48). Clinical complexity was present in 37.2% of trajectories. The latent course distribution was 37.0% gradual, 29.0% rapid, 22.1% delayed, and 11.9% limited response. These quantities are shown in Table 1.

The population was intentionally heterogeneous rather than dominated by a single average curve. Figure 1 shows that this matters for an assessment after four sessions. Rapid responders had already separated from the other

Table 1. Synthetic cohort characteristics

Characteristic	Value
Participants	12,000
Initial symptom severity, mean (SD)	50.94 (9.35)
Initial functioning, mean (SD)	37.17 (8.46)
Chronicity, years, mean (SD)	5.23 (3.48)
Clinical complexity, n (%)	4,465 (37.2)
Rapid response, n (%)	3,483 (29.0)
Gradual response, n (%)	4,443 (37.0)
Delayed response, n (%)	2,646 (22.1)
Limited response, n (%)	1,428 (11.9)

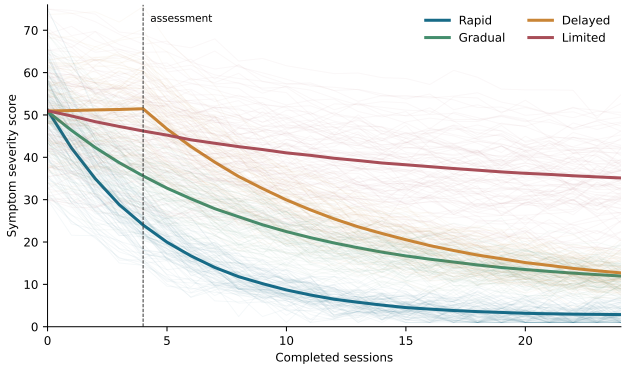

Figure 1. Symptom courses by response pattern.

groups at that time point, whereas delayed responders still resembled limited responders despite a much larger later decline. A policy based only on a fixed initial severity threshold would miss this distinction. Conversely, a rule that stopped every early non-responder would prematurely shorten treatment for a substantial portion of the delayed-response group. The need to distinguish delayed from limited response is consistent with the clinical observation that early progress is informative but not definitive.

The session-8 recovery model had an AUC of 0.965 and the session-16 recovery model had an AUC of 0.766 in the evaluation subset. The notably higher value for the session-8 model is expected: the session-4 proportional change is closely linked to recovery only four sessions later. Discrimination decreased for session-16 recovery because later improvement, especially among delayed responders, is less fully captured by the early assessment. That reduction is clinically desirable to acknowledge. It prevents the allocation model from being interpreted as a diagnostic test with stable accuracy at all time horizons.

The allocation distribution reflected the asymmetric thresholds. Of 6,000 evaluation participants, 30.1% were assigned to 8 sessions, 69.9% to 16 sessions, and one participant (0.02%) to 24 sessions. Figure 2 displays a subsample ordered by proportional session-4 improvement. The coloured field records symptom course rather than a simplified treatment pathway. Endpoint marks show that the rule concentrated 8-session assignments among

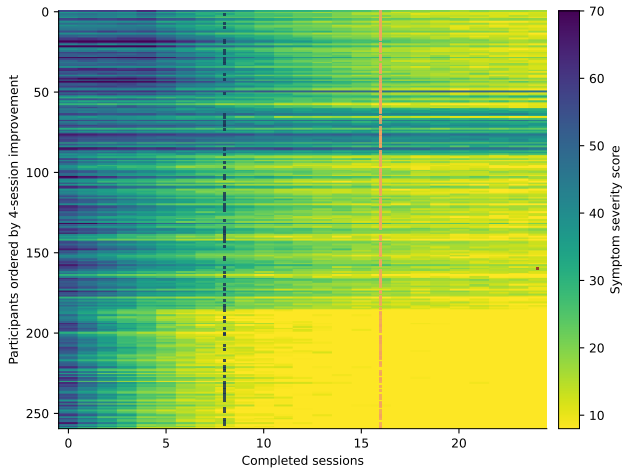

Figure 2. Individual courses and selected duration.

early responders, while nearly all remaining participants were retained to session 16. The single 24-session assignment arose only when the predicted probability of recovery by session 16 was below the prespecified threshold.

This allocation pattern deserves a cautious interpretation. The rule saves appointments mainly by identifying the rapid-response subgroup; it does not claim that delayed response can be safely recognized with high confidence after four sessions. Indeed, the rule places most participants in the 16-session course. Such restraint is important because premature truncation has a different clinical implication from offering a few additional appointments to a person who could have recovered earlier.

3.2. Recovery and session burden

Performance under the four policies is summarized in Table 2. Fixed 8 sessions produced composite recovery in 41.1% of participants (95% bootstrap interval 39.9–42.3). Fixed 16 sessions increased recovery to 86.9% (86.0–87.7), while fixed 24 sessions achieved 87.9% (87.0–88.7). The marginal gain from extending all participants from 16 to 24 sessions was therefore 1.0 percentage point in this cohort. It came at the cost of eight additional sessions for every participant.

Early-response allocation achieved 86.0% recovery (85.1–86.8) with a mean of 13.59 sessions (13.50–13.68). Its recovery rate was 0.90 percentage points lower than fixed 16 sessions, and its mean session use was 2.41 sessions lower. Across 6,000 treatment episodes, the difference corresponds to 14,456 sessions not used by the allocation policy. The final symptom score of 18.28 and functional gain of 12.20 were less favourable than those under fixed 16 sessions, as expected from the shorter average course, but substantially more favourable than those under fixed 8 sessions. These results answer the research question within the synthetic cohort: the four-session rule retained a recovery rate close to the 16-session course while reducing session use by about 15%.

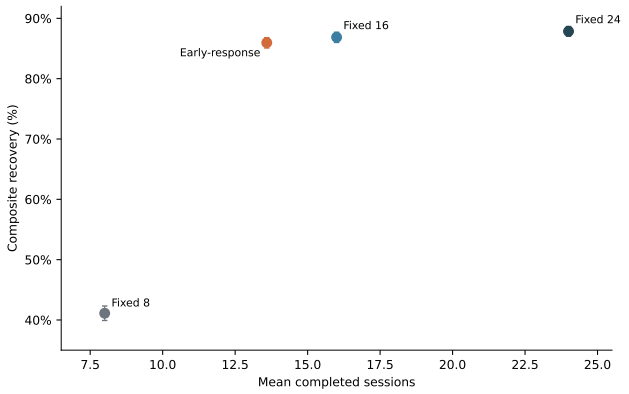

Figure 3. Recovery and completed sessions.

The recovery–capacity relation in Figure 3 displays the same results. The visual position of early-response allocation is informative. It lies close to fixed 16 sessions in recovery but clearly to its left on the completed-session axis. Fixed 24 sessions is slightly higher in recovery, but the horizontal distance illustrates the large service increment required for that small gain. Fixed 8 sessions is efficient only in the narrow sense of appointment use; it has a much lower recovery rate and therefore cannot be considered an interchangeable alternative.

The finding should not be read as a recommendation to shorten all psychotherapy to approximately 14 sessions. The mean is an aggregate consequence of three assignments, and no individual is assigned 13.59 sessions. The operational result is instead that some rapid responders can complete an 8-session course in the synthetic setting without reducing the aggregate recovery rate greatly, while most other participants remain in care until session 16. This distinction is especially important when services convert population averages into individual treatment limits.

The comparison also clarifies why a broad short-versus-long contrast can obscure decision-relevant information. In the generated cohort, an 8-session course is inadequate for many gradual and delayed responders, even though it is sufficient for part of the rapid-response group. A 24-session course is seldom selected by the rule because its threshold design assumes that a 16-session probability estimate can usually identify sufficient recovery. The allocation policy therefore embodies a practical asymmetry: it requires strong early evidence before selecting the shortest course, but it does not routinely extend care to 24 sessions in the absence of a low predicted 16-session probability.

3.3. Sensitivity to delayed response and complexity

The stress tests examined whether the advantage of the allocation rule persisted when delayed response was more common and when complexity more strongly reduced symptom improvement. Figure 4 reports the recovery difference from fixed 16 sessions. Across the 25

Table 2. Policy performance in the evaluation subset

Policy	Recovery, %	Sessions	Final symptom score	Function gain
Fixed 8	41.12	8.00	26.07	9.20
Fixed 16	86.87	16.00	15.96	13.13
Fixed 24	87.85	24.00	12.36	14.56
Early-response allocation	85.97	13.59	18.28	12.20

Recovery requires a symptom score of 30 or lower, a symptom reduction of at least 50%, and an 8-point functional gain. All values are means except recovery.

Figure 4. Recovery difference in stress tests.

settings, early-response allocation was between 0.44 and 1.46 percentage points below fixed 16-session recovery. The corresponding session saving ranged from 0.96 to 3.54 sessions per participant. Recovery remained close to fixed 16 sessions throughout the examined range, but the margin narrowed as delayed response became more common.

The most demanding condition combined a 36% delayed-response proportion with the strongest complexity penalty. Under that condition, the allocation policy was 1.04 percentage points below fixed 16-session recovery and saved 0.96 sessions per participant. The smallest recovery difference occurred with a 12% delayed-response proportion and the strongest complexity penalty, where the difference was 0.44 percentage points and the session saving was 2.88 sessions. The patterns show that early-response allocation is not insensitive to the composition of a service population. Where delayed improvement is common, a conservative policy will correctly retain more participants, reducing the opportunities for efficiency gains. That is an appropriate trade-off rather than a failure of the rule.

These sensitivity findings also highlight a limitation of using only symptom change at a single early point. The delayed-response curve was deliberately structured to have little improvement by session 4. If a real service contained many patients with this pattern, a rule based solely on symptom change could misclassify them. Clinical complexity, treatment preference, attendance difficulty, therapeutic alliance, safety concerns, and diagnostic clarification may add important information. The current model includes complexity as an initial predictor, but it cannot represent the full clinical texture of a psychotherapy encounter. Its value is to make the consequences of a limited predictor set visible.

3.4. Relation to duration evidence and implications

Several duration trials show that treatment format, frequency, and clinical population alter the relation between appointments and outcome. Brief cognitive therapy has produced encouraging results in some panic-disorder trials,[23,24] while other work has found that a brief course can underperform a longer comparison in particular settings.[9] For post-traumatic stress disorder, intensive delivery may produce outcomes comparable with less concentrated schedules for selected populations.[10,11] In depression, once- versus twice-weekly sessions and internet-delivered support have also changed the time structure of treatment without reducing the issue to a single count of appointments.[13,19] These studies support the premise that treatment length should not be separated from delivery pattern and patient characteristics.

The present work extends that premise from a different angle. It does not compare named psychotherapies or claim a superior dose. It tests whether a modest early assessment can select from a small set of fixed course lengths. The result is intentionally parsimonious: the 8-session route requires high predicted recovery, the 16-session route is the default for most patients, and the 24-session route acts as a safeguard for a small group with low predicted recovery. This structure may be more implementable than continuously varying the number of sessions, but it also requires prospective evaluation before use in practice.

Any clinical application should include safeguards. First, an allocation rule should be an aid to shared decision making, not a unilateral discharge mechanism. Patients who prefer further treatment, show emergent risk, or experience functional impairment not captured by a symptom scale should be reviewed individually. Second, outcome measurement must be reliable and timely; a four-session score cannot guide care if it is missing or collected inconsistently. Third, the rule should be recalibrated in the service where it is used because case

mix and treatment delivery influence prediction performance. Poor calibration can create apparently reasonable probability estimates that systematically over- or under-estimate recovery.[38] Finally, any implementation study should compare the allocation policy with usual care using prespecified safety, functioning, and patient-experience outcomes rather than session counts alone.

3.5. Strengths and limitations

The main strength is transparent control of the treatment-duration counterfactual. Each generated participant has a coherent potential course through 24 sessions, allowing direct comparison of policies while holding the underlying response process constant. The model separates development and evaluation subsets, uses a fixed random seed, specifies the equations and thresholds, provides code and generated records, and reports bootstrap intervals. It also treats functional gain as a component of recovery, reducing the risk that appointment saving is portrayed as success when daily functioning has not improved.

Several limitations are equally important. First, the numbers are conditional on the specified equations. The high session-8 AUC reflects the fact that early symptom change is necessarily related to the near-term outcome within the generating process; it should not be expected in a clinical cohort without testing. Second, treatment effects are represented as symptom and functioning paths, not as therapist-patient interactions. The model does not account for treatment switching, medication changes, therapist effects, missed appointments, adverse events, financial barriers, or preference-sensitive decisions. Third, the 24-session route was selected rarely. That result reflects the two thresholds and the simulation parameters; it does not imply that extended psychotherapy is rarely indicated. Fourth, the composite recovery definition contains an 8-point functional gain that is useful for the simulation but requires clinical validation on actual measures. Fifth, the stress tests vary only two features. They do not exhaust the range of possible response distributions or service conditions.

The limitations point directly to the next research step. A prospective study should enroll adults beginning a specified psychotherapy in routine outpatient care, measure symptoms and functioning at every session, fit the allocation model in one set of clinics, and evaluate calibration and safety in different clinics. A randomized implementation design could compare early-response allocation with ordinary duration planning while retaining clinician override. Such a study should report treatment completion, adverse clinical events, functional change, patient preference, and resource use. Model development should follow current prediction-model guidance and use an adequately sized evaluation sample.[35,39] Only that type of study can determine whether the synthetic operating profile is clinically useful.

4. CONCLUSION

This study asked whether a four-session allocation rule could retain recovery close to a fixed 16-session psychotherapy course while reducing completed sessions. Within a fully synthetic cohort of heterogeneous response patterns, the answer was yes: early-response allocation achieved 86.0% composite recovery with 13.6 completed sessions on average, compared with 86.9% recovery and 16 sessions for the fixed 16-session course. The result arose because the rule identified a rapid-response subgroup suitable for an 8-session course while retaining most other participants to 16 sessions. It does not establish that the same policy is effective, safe, or acceptable in clinical practice. Its contribution is a reproducible quantitative rationale for a prospective trial in which early symptom change, functioning, clinician judgment, and patient preference are evaluated together.

BIBLIOGRAFIA / REFERENCES

1. H. U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jönsson, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21(9):655–679, 2011.
2. J. Olesen, A. Gustavsson, M. Svensson, H. U. Wittchen, B. Jönsson, and the CDBE2010 Study Group. The economic cost of brain disorders in Europe. *European Journal of Neurology*, 19(1):155–162, 2012.
3. K. I. Howard, S. M. Kopta, M. S. Krause, and D. E. Orlinsky. The dose-effect relationship in psychotherapy. *American Psychologist*, 41(2):159–164, 1986.
4. P. Cuijpers, M. Huibers, D. D. Ebert, S. L. Koole, and G. Andersson. How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders*, 149(1–3):1–13, 2013.
5. L. Robinson, J. Delgadillo, and S. Kellett. The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychotherapy Research*, 30(1):79–96, 2020.
6. C. Bone, J. Delgadillo, and M. Barkham. A systematic review and meta-analysis of the good-enough level literature. *Journal of Counseling Psychology*, 68(2):219–231, 2021.
7. D. M. Clark, P. M. Salkovskis, A. Hackmann, A. Wells, J. Ludgate, and M. Gelder. Brief cognitive therapy for panic disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 67(4):583–589, 1999.
8. P. Roberge, A. Marchand, D. Reinharz, and P. Savard. Cognitive-behavioral treatment for panic disorder with agoraphobia: A randomized, controlled trial and cost-effectiveness analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(4):475–492, 2008.
9. J. A. Kenardy, M. G. Dow, D. W. Johnston, M. G. Newman, A. Thomson, and C. B. Taylor. A comparison of delivery methods of cognitive-behavioral ther-

apy for panic disorder: An international multicenter trial. *Journal of Consulting and Clinical Psychology*, 71(6):1068–1075, 2003.

10. A. Ehlers, A. Hackmann, N. Grey, J. Wild, S. Liness, I. Albert, et al. A randomized controlled trial of 7-day intensive and standard weekly cognitive therapy for PTSD and emotion-focused supportive therapy. *American Journal of Psychiatry*, 171(3):294–304, 2014.
11. E. B. Foa, C. P. McLean, Y. Zang, D. Rosenfield, E. Yadin, J. S. Yarvis, et al. Effect of prolonged exposure therapy delivered over 2 weeks vs 8 weeks vs present-centered therapy on PTSD symptom severity in military personnel: A randomized clinical trial. *JAMA*, 319(4):354–364, 2018.
12. H. Christensen, K. M. Griffiths, A. Mackinnon, and K. Brittliffe. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. *Psychological Medicine*, 36(12):1737–1746, 2006.
13. H. Hadjistavropoulos, V. Peynenburg, D. Thiessen, M. Nugent, E. Karin, B. Dear, et al. A randomized factorial trial of internet-delivered cognitive behavioural therapy: An 8-week program with or without extended support and booster lesson. *Internet Interventions*, 27:100499, 2022.
14. P. Knekt, O. Lindfors, T. Härkänen, M. Valikoski, E. Virtala, M. A. Laaksonen, et al. Randomized trial on the effectiveness of long- and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychological Medicine*, 38(5):689–703, 2008.
15. S. Lorentzen, T. Ruud, A. Fjeldstad, and P. Høglend. Comparison of short- and long-term dynamic group psychotherapy: Randomised clinical trial. *British Journal of Psychiatry*, 203(3):280–287, 2013.
16. M. Barkham, A. Rees, D. A. Shapiro, W. B. Stiles, R. M. Agnew, J. Halstead, et al. Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology*, 64(5):1079–1085, 1996.
17. D. A. Shapiro, M. Barkham, A. Rees, G. E. Hardy, S. Reynolds, and M. Startup. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, 62(3):522–534, 1994.
18. J. Dekker, P. J. Molenaar, S. Kool, G. van Aalst, J. Peen, and F. de Jonghe. Dose-effect relations in time-limited combined psycho-pharmacological treatment for depression. *Psychological Medicine*, 35(1):47–58, 2005.
19. S. J. E. Bruijniks, L. Lemmens, S. D. Hollon, F. Peeters, P. Cuijpers, A. Arntz, et al. The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. *British Journal of Psychiatry*, 216(4):222–230, 2020.
20. S. F. McMain, A. L. Chapman, J. R. Kuo, K. L. Dixon-Gordon, T. H. Guimond, C. Labrish, et al. The effectiveness of 6 versus 12 months of dialectical behavior therapy for borderline personality disorder: A noninferiority randomized controlled trial. *Psychotherapy and Psychosomatics*, 91(6):382–397, 2022.
21. N. Nacasch, J. D. Huppert, Y.-J. Su, Y. Kivity, Y. Dinshtein, R. Yeh, et al. Are 60-minute prolonged exposure sessions with 20-minute imaginal exposure to traumatic memories sufficient to successfully treat PTSD? A randomized noninferiority clinical trial. *Behavior Therapy*, 46(3):328–341, 2015.
22. L. Dell, A. M. Sbisa, A. Forbes, M. O'Donnell, R. Bryant, S. Hodson, et al. Effect of massed versus standard prolonged exposure therapy on PTSD in military personnel and veterans: A non-inferiority randomised controlled trial. *Psychological Medicine*, 53(4):1323–1332, 2023.
23. J. D. Herbert, A. A. Rheingold, B. A. Gaudiano, and V. H. Myers. Standard versus extended cognitive behavior therapy for social anxiety disorder: A randomized-controlled trial. *Behavioural and Cognitive Psychotherapy*, 32(2):131–147, 2004.
24. M. Böhni, H. Spindler, M. Arendt, E. Hougaard, and N. Rosenberg. A randomized study of a massed three-week cognitive behavioural therapy schedule for panic disorder. *Acta Psychiatrica Scandinavica*, 120(3):187–195, 2009.
25. M. Böttche, B. Wagner, M. Vöhringer, M. Heinrich, J. Stein, P. Selmo, et al. Is only one cognitive technique also effective? Results from a randomized controlled trial of two different versions of an internet-based cognitive behavioural intervention for post-traumatic stress disorder in Arabic-speaking countries. *European Journal of Psychotraumatology*, 12(1):1943870, 2021.
26. M. Kool, H. L. van, A. Bartak, S. C. M. de Maat, A. Arntz, J. W. van den Eshof, et al. Optimizing psychotherapy dosage for comorbid depression and personality disorders: A pragmatic randomized factorial trial using schema therapy and short-term psychodynamic psychotherapy. *BMC Psychiatry*, 18:252, 2018.
27. O. J. Storebø, J. M. Stoffers-Winterling, B. A. Völlm, M. T. Kongerslev, J. T. Mattivi, and M. S. Jørgensen. Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, 5:CD012955, 2020.
28. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10:89, 2021.
29. G. H. Guyatt, A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650):924–926, 2008.
30. S. Juul, J. C. Jakobsen, C. K. Jørgensen, S. Poulsen, P. Sørensen, and S. Simonsen. The difference between shorter- versus longer-term psychotherapy for adult mental health disorders: A systematic review with meta-analysis. *BMC Psychiatry*, 23:438, 2023. https://doi.org/10.1186/s12888-023-04895-6
31. J. Delgadillo, O. Moreea, and W. Lutz. Different people respond differently to therapy: A demonstration

using patient profiling and risk stratification. *Behaviour Research and Therapy*, 79:15–22, 2016.

32. J. Delgadillo, D. Huey, H. Bennett, and D. McMillan. Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, 85(9):835–853, 2017.
33. A. Burton, D. G. Altman, P. Royston, and R. L. Holder. The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292, 2006.
34. T. P. Morris, I. R. White, and M. J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102, 2019.
35. G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1):55–63, 2015.
36. K. G. M. Moons, D. G. Altman, J. B. Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W. Steyerberg, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, 162(1):W1–W73, 2015.
37. R. F. Wolff, K. G. M. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, 170(1):51–58, 2019.
38. B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg. Calibration: The Achilles heel of predictive analytics. *BMC Medicine*, 17:230, 2019.
39. R. D. Riley, G. M. Snell, Y. Ensor, A. J. Burke, F. E. Harrell, K. G. M. Moons, and G. S. Collins. Minimum sample size for developing a multivariable prediction model: PART II–binary and time-to-event outcomes. *Statistics in Medicine*, 38(7):1276–1296, 2019.
40. P. Royston, D. G. Altman, and R. Sauerbrei. Discrimination and calibration for a risk prediction model with a binary outcome. *Statistics in Medicine*, 28(29):3647–3670, 2009.