Article:

Napoleone, E. orcid.org/0000-0003-2172-5345, Evans, C. orcid.org/0000-0002-4197-4202, Patalay, P. orcid.org/0000-0002-5341-3461 et al. (2 more authors) (2019) Trajectories of change of youth depressive symptoms in routine care: shape, predictors, and service-use implications. European Child and Adolescent Psychiatry, 28 (11). pp. 1527-1536. ISSN 1018-8827

https://doi.org/10.1007/s00787-019-01317-5

Trajectories of Change of Youth Depressive Symptoms in Routine Care: Shape, Predictors, and Service-Use Implications

Authors:

Napoleone, Elisa[1]

Evans, Chris[2]

Patalay, Praveetha[3]

Edbrooke-Childs, Julian[1]

Wolpert, Miranda[1]

[1] Anna Freud National Centre for Children and Families and University College London, 12 Maresfield Gardens, London, NW3 5SU, United Kingdom

[2] University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

[3] University of Liverpool, Foundation Building, Brownlow Hill, Liverpool, L69 7ZX, United Kingdom

Running Head: Youth Depression Trajectories in Routine Care

Abstract

Objective: Depression is one of the main reasons for youth accessing mental health services, yet we know little about how symptoms change once youth are in routine care. This study used multilevel modelling to examine the average trajectory of change and the factors associated with change in depressive symptoms in a large sample of youth seen in routine mental health care services in England.

Method: Participants were 2,336 youth aged 8 to 18 (mean age = 14.52; 77% females; 88% white ethnic background) who tracked depressive symptoms over a period of up to 32 weeks while in contact with mental health services. Explanatory variables were age, gender, whether the case was closed, total length of contact with services, and baseline severity in depression scores.

Results: Faster rates of improvement were found in older adolescents, males, those with shorter time in contact with services, closed cases, and those with more severe symptoms at baseline.

Conclusions: This study demonstrates that, when youth self-report their depressive symptoms during psychotherapy, symptoms decrease in a linear trajectory. Attention should be paid to younger people, females, and those with lower than average baseline scores, as their symptoms decrease at a slower pace compared to others.

Key words: Trajectories of change; youth depressive symptoms; self-report; routine data; mental health

One of the challenges in routine psychotherapy delivery is to ensure that clients receive effective and efficient care, i.e. treatment that is responsive to individual needs and results in improvement, while preventing unnecessary resource use. Understanding the way in which clients' symptoms change during routine psychological treatment offers the potential to tackle this challenge [1], and has been the subject of considerable research efforts with adult populations [2-5], but to a far lesser degree with children and adolescents.

The trajectory of change that has been identified in the adult psychological therapy literature most frequently is a curvilinear one, where change occurs rapidly at first, and then decelerates [6,7]. It has been proposed that the shape varies as a function of the overall treatment length, with clients in shorter treatment spells changing in a linear fashion until they reach a good-enough level to end therapy [8]. With routine outcome monitoring (i.e., patients reporting on their own symptoms on a regular basis, e.g., at each session) becoming increasingly common in mental health services [9], data collected at frequent intervals offers rich information to put these models to the test.

Examining change at multiple points during treatment is more informative than traditional pre-post treatment measurement approaches, as one can make full use of the wealth of information that is otherwise lost when only two discrete time points are used for analyses [10]. In addition, statistical advances in recent years allow for sophisticated analyses that take into account all available data and are able to estimate individual rates of change over time through hierarchical modelling, achieving great precision and power [11].

These methods have started being used to explore trajectories of change in youth routine mental health care [12-15]. This research group used a global functioning measure (the Youth Outcomes Questionnaire [16]) to model curvilinear (log-linear or square-root) change as a function of either session number or weeks in treatment. A number of individual characteristics, including severity at baseline, age and gender, were used to predict different trajectories. On the basis of the overall trajectory models, a warning system was developed to alert practitioners of cases at risk of treatment failure, and it achieved reasonable accuracy, leading the authors to argue for its use in routine practice.

These studies highlight the practical implications that examining trajectories of change can have. However, their findings can be extended on a number of fronts. First, using a measure of global functioning may not detect change that occurs in specific symptoms with which youth present in routine care. Second, there are questions regarding the generalizability of the results in terms of healthcare context and sample characteristics. For instance, Nelson and colleagues' study [15] comprised data from a large privately-managed care organization, in which fewer than 10% of the children had multiple diagnoses. This is arguably very different from the type of children who access the publicly-funded health service, such as in the United Kingdom (UK), where comorbidities are higher and possibly due to limited resources, severity may be greater [17,18].

Depression is a common presenting difficulty in routine care, and mounting evidence suggests that outcomes for depressed youth seen in these settings have little resemblance to the effectiveness rates reported for randomized controlled trials [19,20]. This, alongside evidence that depressive symptoms at a young age predict psychopathology later in life [21] and put youth at increased risk of poor psychological outcomes [22], underscores the importance of understanding trajectories and predictors of symptom change of this disorder in routine care.

The present research addresses this gap by examining the depressive symptom change trajectories in a large naturalistic sample of children and young people seen in routine care by mental health services in the UK. The aims were: a) to describe the average trajectory of change that depressive symptoms take in youth aged 8 to 18 years old while in contact with services, and b) to investigate variance in trajectories in relation to demographic (age and gender), service use (case closure status, length of contact, whether there had been any prior contact) and severity (baseline depressive symptoms).

Method

Participants and Procedure

This study used data from a database of children and young people seen as part of the Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) service transformation program in England between 2011 and 2015 [23]. Data were collected from 81 Child and Adolescent Mental Health Services (CAMHS) within the National Health Service (NHS), local authorities, and voluntary sector providers[18]. Key aspects of the programme were practitioner training both in evidence-based interventions (e.g., CBT and interpersonal therapy for depression) and in use of routine outcome monitoring. A sample of 2,336 children and young people between the ages of 8 and 18 who had completed a self-reported depression symptom measure on at least three occasions during the episode of care, and for whom there was information about their gender, was selected for analysis. Cases in this sample had a mean age of 14.52 (*SD* = 1.72) and comprised 1,803 (77%) females. Symptoms were tracked for an average of 14.44 weeks (*SD* = 7.21; *Median* = 13.86; *Range* = 2-32), with a total of 10,925 sessions in which a depression score was recorded, and an average of 25.92 days (*SD* = 14.85; *Median* = 22; *Range* = 6-86) between sessions. According to clinician report, available for 51% (1,182), the most prevalent presenting problem was "Depression/low mood" (1,043; 88%). In addition many reported a range of anxiety problems including "Anxious in social situations" (784; 66%) and "Anxious generally" (718; 61%; categories are not mutually exclusive). Clinician report of type of therapy was only available for 48% (1,121 cases). The most common type of therapy received in the sessions where depression was also tracked was Cognitive Behavioral therapy (787; 70%), followed by Multimodal Integrated therapy (23%) and Family Systemic therapy (18%; note that categories are not mutually exclusive).

Due to the naturalistic setup of the original database, we undertook the procedures detailed below to obtain the analysis sample described above. A case was defined as a child or young person between the ages of 8 and 18 seen for an episode of care within a service. Overlapping episodes (due to data errors or because the case was seen by multiple teams) were merged into one. A session was defined as a day in which one or more events were recorded for a case; where there were records apparently for multiple contacts within the same day, information from the entry with the most complete information was used, or where information was conflicting (e.g. there was both a face-to-face and a non-therapeutic contact recorded on the same day), the entry relating to the therapeutic contact was used in the analyses. This resulted in 8,238 episodes of care and 26,814 sessions in which a depression symptom score was recorded. Some cases were marked as open at the most recent point of data reporting and this variable was used to check on potential trajectory censorship.

To be able to model the shape of change over time, and to maximize the chance of including cases that had chosen to track depressive symptoms, we selected cases who filled out the depression measure on at least three occasions (n = 3,123). Similarly to other studies [24,15,25], if two consecutive sessions within an episode of care were more than 90 days apart, we took the last session before the gap as the end of that episode, and the first session after the gap as the start of a new episode; only the episode with most sessions was retained to preserve independence of observations (a total of 84% first episodes, 14% second episodes, 2% third episodes, and less than 1% fourth episodes; n = 2,601). Finally, in line with previous research [12,14], we excluded weeks in treatment beyond the 90th percentile, which in our sample was 32 weeks since the first session in which a depression symptom score was recorded.[1] The selected sample therefore comprised 2,336 cases (28% of the ‘depressive symptoms’ sample). Figure 1 provides a graphical description of the sample selection process.

< Insert Figure 1 about here >

[1] We conducted a sensitivity analysis on the uncensored sample. The results were broadly similar, the only notable difference being the fixed main effect of total length of contact being significant in the uncensored sample, which we attribute to the larger variability in this variable that followed from including the top 10% of contact lengths.

Table 1 displays the key characteristics of the eligible, and the selected samples. The samples were mostly similar, with the selected sample including a slightly higher proportion of females and white ethnic backgrounds and higher severity. They had longer contact with the service which is to be expected given the selection criteria. The selected sample was similar to the most recently published figures [26,27] on prevalence of mental health disorders and caseload characteristics of services in England in terms of ethnicity (between 87% and 89% white, here 88%), but not in terms of gender (between 41% and 49% girls, here 77%).

< Insert Table 1 about here >

No ethical review was required as this study involved secondary analysis of routinely collected anonymized data.

Measures

Service use. Length of contact in weeks was derived from the dates of first and last sessions in which a depression measure was recorded. Information on whether a case had prior contact with the service was derived through the procedure to identify episodes with data no more than 90 days apart (described in the Participants and Procedure section).

Severity in depressive symptoms. The Major Depressive Disorder (MDD) subscale of the Revised Child Anxiety and Depression Scale (RCADS [28]) was used to measure self-reported depressive symptoms. The MDD subscale comprises 10 items rated from Never (0) to Always (3), for a maximum total severity score of 30. A cut-off score of 11 is reported to achieve adequate sensitivity (74%) and specificity (77%) to distinguish between normative and clinical samples [29]. The full RCADS comprises five more subscales that measure different types of anxiety. Common practice in the services that took part in this study was to administer the full RCADS at assessment, and then choose one or more subscales with which to track symptomatic change during treatment [30]. The depression subscale of the RCADS has shown adequate psychometric properties in other samples (alpha = .87 and .76, [29,28]); in the present study, Cronbach's alpha for the depression subscale at baseline was .85 (n = 2,336).

Statistical Analyses

Multilevel modelling was used, this analyses *individual* growth curves while accounting for dependency inherent in repeated measures data. Hence, measurement occasions (Level I) were treated as nested within individuals (Level II). Level III (services) only accounted for 3% of the variance in depression scores and was therefore excluded from the models to minimize complexity. Weeks since the first session in which a depression score was recorded was used to order scores over time and predict change in outcome, in line with previous research [12,13]. To be able to include baseline depression scores as a predictor, models were fitted on data that excluded the baseline scores (i.e. in the models 'time' = 0 corresponds to the first session after baseline assessment).[2] To model curvilinear trajectories, the time variable was transformed to log-linear and square-root. Polynomials (e.g., quadratic) were also considered, but given that baseline scores were included as an explanatory variable, not all cases had enough measurement occasions to accurately estimate a curve in this way (e.g. cases with three sessions only).

Data were analyzed using a maximum-likelihood estimation procedure, which estimates all model parameters simultaneously to maximize the likelihood that the estimates of effects are representative of the population effects. The maximum-likelihood approach is the appropriate method for comparing nested models, not only when they differ in their random but also in their fixed parts [31]. Model building was approached with a stepwise strategy to test the predictor variables (see above). Nested models (i.e., models that share the same parameters [32]) were compared using the deviance (likelihood ratio) test, alongside the Akaike Information Criterion (AIC). A significant deviance test and a lower AIC value, indicate an improvement in the model. All analyses were performed using R version 3.3.2 [33]. The models were built using the package nlme version 3.1-128 [34]. Effect size was calculated by dividing the average difference between baseline and last scores by the standard deviation at baseline [35]. This approach doesn't adjust change by baseline score and provides a conservative estimate of the effect size.

Results

Descriptive Data: Change and Clinical Significance

The average score of depressive symptoms at the first time point was 17.67 (SD = 5.75), and at the last time point it was 13.6 (SD = 7.28), yielding an effect size of d = 0.71 (95% CI 0.66, 0.75). Of the cases that were above the cut-off at the first time point (2,072), 616 (30%) were no longer above the cut-off at a last time point ("recovered"). Looking at whether the change from a first to a last session was greater than what would be expected by measurement error alone (greater than 6.2, calculated using the "reliable change criteria" [36]), 742 cases (32%) reliably improved, 1,501 cases (64%) did not reliably change, and 93 (4%) reliably deteriorated. Combining the two criteria, 474 cases reliably "recovered" (23% of those above the cut-off at the first time point). These rates are in line with those reported in other routine mental health samples internationally [37].

Trajectories of Change

[2] Model building was also approached including all scores from baseline in the outcome variable, and the results were similar.

Model building. Table 2 summarizes the model building steps, and the fit parameters (AIC and -2 Log Likelihood) at each step, as well as significance testing of the nested models (see table footnote). As can be seen, the unconditional linear growth model with both fixed and random slope (Model 2b) fitted the data slightly better based on the AIC values than both the log-linear one (Model 2d) and the square-root one (Model 2f). Therefore, Model 2b was used as the basis for including the explanatory variables. Models 3 and 4 included explanatory variables and their interactions with time (rates of change), with the final model being more parsimonious and retaining only the significant explanatory variables. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

< Insert Table 2 about here >

Final model. The final model is a model that included a random linear intercept and slope of time in weeks, and demographic, service use, and severity characteristics as explanatory variables (see Table 3 for fixed and random effects' estimates). *C*ompared to the unconditional models, this model accounted for 59% of the between-subjects variance of the intercept[3], 5% of the between-subjects variance in slope[4] and 40% of all the within-subjects variance in depression scores[5].

< Insert Table 3 about here >

The intercept estimate shows that cases had an average depression score of 16 after baseline, and that each week, this score was estimated to decrease by 0.3 points overall. As scores decreased overall, positive slopes (interactions with time) are to be interpreted as slower rates of change. Children and young people who were older than the average in this sample (15 years old), compared to younger children were estimated to have a higher post-baseline score and a faster rate of change. Girls, compared with boys, started with higher scores, but their rate of improvement was slower. Cases that were closed had a lower post-baseline score, and a faster rate of improvement compared to open cases. Cases with longer than average total length of contact (17 weeks) had average post-baseline scores but slower improvement compared to cases that stayed in contact with the service for fewer weeks. Finally, those with higher depressive symptoms at baseline also had a higher post-baseline score, and their rate of change was faster than those with lower baselines. Figure 2 provides a representation of the estimated effects of each of the explanatory variables on the linear trajectory of change in depression symptom scores over the course of

[3] Pseudo R^2 = 1 – (Final Model Intercept Variance/Model 2b Intercept Variance)

[4] Pseudo R^2 = 1 – (Final Model Slope Variance/Model 2b Slope Variance)

[5] Pseudo R^2 = 1 – (Final Model Within-subjects Residual/Model 1 Within-subjects Residual)

31 weeks after the first recorded score. For continuous variables, the mean, as well as one or two standard deviations from the mean, are plotted. The horizontal reference line indicates the point below which scores are no longer considered to belong to a clinical sample (i.e., 11 [29]).

< Insert Figure 2 about here >

Discussion

The present study sought to explore trajectories of change in depressive symptoms of children and adolescents seen in routine mental health services in the UK. We found that improvement was best described as a steady linear trajectory over time. Rates of change in scores over time depended on demographic, severity, and service use factors. Faster rates of improvement (i.e., steeper average slopes) were found in older adolescents, males, cases that had spent a shorter time in contact with services, closed cases, and those with higher baseline depression scores. Differences by each predictor were statistically significant but not large and the model predicted only 5% of the between individuals variance in rate of change but 40% of the total variance in scores against time.

To the best of our knowledge, this is the first study to explore depressive trajectories in a naturalistic sample of youth seen in routine care. Unlike previous research (e.g., [12,15]) from different settings and using different measures, we did not find evidence for a meaningful curvilinear trajectory of change over time with both log-linear and square-root transformations of weeks in contact with a service fitting slightly worse than the simpler linear model. This could be due to differences in sample characteristics and in the context in which the youth received treatment. For instance, Nelson et al.'s study [15] was conducted in a private managed care organization, while our sample consisted of services taking part in a statutory transformation program to improve access to evidence-based care [18]. These services were under considerable demand and financial strain [23] and this may have resulted in youth being discharged earlier than their counterparts in other studies, preventing us from observing a curvilinear trajectory of change. In addition, Nelson et al.'s sample [15] only comprised 8% with multiple conditions, while the available data in our sample suggests that comorbidities such as social and generalized anxiety were very common (more than 60%). It may be that for less severe and complex cases, an improvement followed by a plateau is a more reasonable trajectory than for more severe cases, where a steady linear decline in symptoms can be observed as in the present study. However, methodological differences may also contribute to the discrepancy. As previous studies only reported AIC/BIC [15,13,14] or square-root [12] models rather than significance tests, comparability is restricted. It may also be that differences in numbers of scores per participant affected power to test differences and the high proportion of short therapies in this sample have that effect. That cannot remove the finding that the AIC values were better for the linear model here than the curvilinear models.

This study tested variables relationships with speed of improvement. For some factors, findings were broadly consistency with the literature e.g., higher baseline scores predict faster improvement [12,38,15], however, for others findings were more discrepant. Firstly, we found that older adolescents had higher baseline scores, but also improved more rapidly than younger ones. This is consistent with findings of Nelson and colleagues [15], but not Cannon and colleagues [12], who found that older youth had lower scores at the start, but no significant differences in rates of change. It may be that adolescents in our study were more receptive to treatment than younger children, or that the treatment they received in routine care was better tailored at their age group and cognitive abilities (e.g. Cognitive Behavioral therapy). Secondly, with regards to gender, females in our study had higher depressive scores at baseline but improved at a slower rate compared to males; this is partially consistent with results for global functioning outcomes in previous studies (Cannon et al. [12] found females to have higher baselines, but no differences in rates of change) but not with others (Nelson et al. [15] found lower baselines for females, and no significant differences in improvement rates, although the non-significant coefficient was in the same direction as in our study). Finally, with regards to the service use factors, while both case closure status and length of contact are indicators that are obviously not available at the start of treatment, and therefore they are not useful as potential guides for treatment selection or setting expectations, they do account for differences in the average trajectory found in this study. As this was a large routine sample with both closed and opened cases, it was important to include these variables to account for these different types of cases. Case closure separates censored (potentially unfinished) trajectories from uncensored ones helping minimize the impact of censorship on estimating the effects of other variables. That closed cases had steeper rates of improvement suggests less complexity for these cases. Similarly, those that were still open by 32 weeks were likely to be more complex and therefore require more resources. In addition, the significant interaction between weeks in contact with the service, and total length of contact is in line with the "good-enough" level hypothesis [8], whereby those who finish treatment earlier improve faster than others, as they leave treatment when their symptoms are at a "good-enough" level.

Our findings need to be considered in light of inevitable limitations. Firstly, with regards to the data, although the large, naturalistic sample provides high statistical power, external validity, and some promise of generalizability, this generalizability is compromised as we cannot know whether youth who tracked their depressive symptoms were different from those who did not. In addition, we were unable to fully characterize the sample in terms of presenting difficulties, comorbidities and type of treatment (both psychological and pharmacological) as this information was only available through clinician report and for a limited portion of the sample. Future studies should endeavor to gather this information more consistently, and incorporate it into the analyses to fine tune models to these important characteristics. We were also unable to consider how other symptoms in addition to depression changed during treatment; As depressive symptoms are often comorbid with anxiety, it will be important for future research to explore interactions between the symptoms throughout treatment. Also, in terms of representativeness, while the ethnic background was similar to nationally-reported rates [26,27] our sample included more females, although this gender imbalance is in line with recently-reported prevalence rates of these symptoms in adolescence [39]. It should also be remembered that there are often differences in the ratings of change when self-report is compared with clinician, family, or teacher ratings. For example, a recent meta-analysis noted that, for depression, self-reported outcomes in youth tend to show more improvement than either parent- or teacher-reported outcomes [40]. Finally, in terms of methodological considerations, the multilevel modelling approach is increasingly widely used and in our data it explained considerable variance in scores. However, we believe that further analyses are needed to explore whether other statistical techniques, such as the application of propensity scores, growth curve mixture models [41], or "nearest neighbor" score predictions [42], provide even better models that can guide our understanding of the way depressive symptoms change for youth that receive treatment in routine care.

As our overall model explained a substantial amount of variation in depression scores, it may be possible, using such analyses, to develop a system whereby trajectories are predicted at the beginning of depression treatment based on factors that are known (e.g. age, gender and baseline severity) and are adjusted as treatment develops based on the rate of change. As discussed, this approach has been used for general functioning outcomes in both adult (e.g., [9]) and youth (e.g., [15]) literature, and has shown promising results. For instance, Boswell and colleagues [43] discuss how systems based on a functional outcome questionnaire are able to predict between 85% and 100% of clients who deteriorate before they leave care, a rate that is much higher than clinician judgement alone. Development of such a system for specific disorder symptoms is needed, as those are often used in clinical practice and found to be more acceptable than general ones, and better able to pick up on the changes that are targeted during care. A simple application through a computerized system, or even through reference to plots such as those of Figure 2, could be used by clinicians to discuss options and expectations with youth and their families. This study presents the first stage of this development process looking at youth depression outcomes, and highlights how understanding the way individuals respond to treatment can have practical implications for guiding their progress through routine mental health care.

Conflict of Interest Statement: The first, fourth and fifth authors were involved in the programme of service transformation that this manuscript draws on. The fifth author led the outcomes and evaluation group that agreed the approach to measurement used in the initiative.

References

1. Saunders R, Cape J, Fearon P, Pilling S (2016) Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. Journal of Affective Disorders 197:107-115. doi:10.1016/j.jad.2016.03.011

2. Lutz W, Rubel J, Schiefele A-K, Zimmermann D, Böhnke JR, Wittmann WW (2015) Feedback and therapist effects in the context of treatment outcome and treatment length. Psychotherapy Research 25 (6):647-660. doi:10.1080/10503307.2015.1053553

3. Owen JJ, Adelson J, Budge S, Kopta SM, Reese RJ (2016) Good-enough level and dose-effect models: Variation among outcomes and therapists. Psychotherapy Research 26 (1):22-30. doi:10.1080/10503307.2014.966346

4. Percevic R, Lambert MJ, Kordy H (2006) What is the predictive value of responses to psychotherapy for its future course? Empirical explorations and consequences for outcome monitoring. Psychotherapy Research 16 (3):364-373. doi:10.1080/10503300500485524

5. Rubel J, Lutz W, Kopta SM, Köck K, Minami T, Zimmermann D, Saunders SM (2015) Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. Psychological Assessment 27 (2):478-488. doi:10.1037/pas0000060

6. Baldwin SA, Berkeljon A, Atkins DC, Olsen JA, Nielsen SL (2009) Rates of change in naturalistic psychotherapy: Contrasting dose–effect and good-enough level models of change. Journal of Consulting and Clinical Psychology 77 (2):203-211. doi:10.1037/a0015235

7. Howard KI, Kopta SM, Krause MS, Orlinsky DE (1986) The dose–effect relationship in psychotherapy. American Psychologist 41 (2):159-164

8. Barkham M, Connell J, Stiles WB, Miles JNV, Margison F, Evans C, Mellor-Clark J (2006) Dose-effect relations and responsive regulation of treatment duration: The good enough level. Journal of Consulting and Clinical Psychology 74 (1):160-167. doi:10.1037/0022-006x.74.1.160

9. De Jong K, Timman R, Hakkaart-Van Roijen L, Vermeulen P, Kooiman K, Passchier J, Busschbach JV (2014) The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. Psychotherapy Research 24 (6):629-639. doi:10.1080/10503307.2013.871079

10. Hayes AM, Feldman GC, Beevers CG, Laurenceau J-P, Cardaciotto L, Lewis-Smith J (2007) Discontinuities and cognitive changes in an exposure-based cognitive therapy for depression. Journal of Consulting and Clinical Psychology 75 (3):409-421. doi:10.1037/0022-006x.75.3.409
11. Singer JD, Willett JB (2003) Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press, Oxford:
12. Cannon JAN, Warren JS, Nelson PL, Burlingame GM (2010) Change Trajectories for the Youth Outcome Questionnaire Self-Report: Identifying Youth at Risk for Treatment Failure. Journal of Clinical Child & Adolescent Psychology 39 (3):289-301. doi:10.1080/15374411003691727
13. Warren JS, Nelson PL, Burlingame GM (2009) Identifying Youth at Risk for Treatment Failure in Outpatient Community Mental Health Services. Journal of Child and Family Studies 18 (6):690-701. doi:10.1007/s10826-009-9275-9
14. Warren JS, Nelson PL, Burlingame GM, Mondragon SA (2012) Predicting patient deterioration in youth mental health services: community mental health vs. managed care settings. Journal of Clinical Psychology 68 (1):24-40. doi:10.1002/jclp.20831
15. Nelson PL, Warren JS, Gleave RL, Burlingame GM (2013) Youth Psychotherapy Change Trajectories and Early Warning System Accuracy in a Managed Care Setting: Youth Outcomes and Warning System Accuracy. Journal of Clinical Psychology 69 (9):880-895. doi:10.1002/jclp.21963
16. Burlingame GM, Wells MG, Lambert MJ, Cox JC (2004) Youth Outcome Questionnaire. In: Maruish M (ed) The use of psychological tests for treatment planning and outcome assessment, vol 2. 3 edn. Erlbaum, Mahwah, NJ, pp 235-274
17. Children’s, Commissioner, for, England (2017) Briefing: Children’s Mental Healthcare in England. Children's Commissioner, London
18. Wolpert M, Jacob J, Napoleone E, Whale A, Calderon A, Edbrooke-Childs J (2016) Child- and Parent-reported Outcomes and Experience from Child and Young People’s Mental Health Services 2011–2015. CAMHS Press, London
19. Bear H, Edbrooke-Childs JH, Norton S, Krause K, Wolpert M (2018) Treatment response following routine mental health care among children and adolescents with anxiety and/or depression: A systematic review and meta-analysis. Submitted
20. Weersing VR, Weisz JR (2002) Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. Journal of Consulting and Clinical Psychology 70 (2):299-310. doi:10.1037//0022-006x.70.2.299
21. Lima NNR, do Nascimento VB, de Carvalho SMF, de Abreu LC, Neto MLR, Brasil AQ, Celestino

Júnior FT, de Oliveira GF, Reis AOA (2013) Childhood depression: a systematic review. Neuropsychiatric Disease and Treatment:1417. doi:10.2147/ndt.s42402
22. Dekker MC, Ferdinand RF, van Lang NDJ, Bongers IL, van der Ende J, Verhulst FC (2007) Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. Journal of Child Psychology and Psychiatry 48 (7):657-666. doi:10.1111/j.1469-7610.2007.01742.x
23. Department of, Health (2015) Future in Mind: Promoting, protecting and improving out children and young people's mental health and wellbeing. Department of Health and NHS England, London
24. Bickman L, Douglas SR, De Andrade ARV, Tomlinson M, Gleacher A, Olin S, Hoagwood K (2016) Implementing a Measurement Feedback System: A Tale of Two Sites. Administration and Policy in Mental Health and Mental Health Services Research 43 (3):410-425. doi:10.1007/s10488-015-0647-8
25. Reese RJ, Toland MD, Hopkins NB (2011) Replicating and extending the good-enough level model of change: Considering session frequency. Psychotherapy Research 21 (5):608-619. doi:10.1080/10503307.2011.598580
26. Green H, McGinnity A, Meltzer H, Ford T, Goodman R (2005) Mental health of children and young people in Great Britain 2004. Palgrave Macmillan, Basingstoke, UK
27. Barnes D, Wistow R, Dean R, Foster B (2006) National child and adolescent mental health service mapping exercise 2005. Durham University, School of Applied Social Sciences, Durham
28. Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE (2000) Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. Behaviour Research and Therapy 38 (8):835-855
29. Chorpita BF, Moffitt CE, Gray J (2005) Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. Behaviour Research and Therapy 43 (3):309-322. doi:10.1016/j.brat.2004.02.004
30. Law D (2012) A practical guide to using service user feedback & outcome tools to inform clinical practice in child & adolescent mental health. Retrieved August 26:2014
31. Tabachnick BG, Fidell LS (2007) Using multivariate statistics. Pearson/Allyn & Bacon, Boston
32. Snijders TAB, Bosker RJ (2012) Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. 2nd edn. Sage Publishers, London
33. R., Core, Team (2016) R: A language and environment for statistical computing. 3.3.2 edn. R Foundation for Statistical Computing, Vienna, Austria
34. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2016) nlme: Linear and Nonlinear Mixed Effects

Models. R package version 3.1-128.
35. Becker BJ (1988) Synthesizing standardized mean-change measures. British Journal of Mathematical and Statistical Psychology 41 (2):257-278
36. Jacobson NS, Truax P (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 59 (1):12
37. Edbrooke-Childs JH, Wolpert M, Zamperoni V, Napoleone E, Bear H (2018) Evaluation of reliable improvement rates in depression and anxiety at the end of treatment in adolescents. BJPsych Open 4 (4):250-255
38. Jackson DS, Keir SS, Sender M, Mueller CW (2017) Reliable Change and Outcome Trajectories Across Levels of Care in a Mental Health System for Youth. Administration and Policy in Mental Health and Mental Health Services Research 44 (1):141-154. doi:10.1007/s10488-015-0690-5
39. Patalay P, Fitzsimons E (2016) Correlates of mental illness and wellbeing in children: are they the same? Results from the UK Millennium Cohort Study. Journal of the American Academy of Child & Adolescent Psychiatry 55 (9):771-783
40. Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, Jensen-Doss A, Hawley KM, Krumholz Marchette LS, Chu BC, Weersing VR, Fordwood SR (2017) What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. American Psychologist 72 (2):79-117. doi:10.1037/a0040360
41. Queen AH, Barlow DH, Ehrenreich-May J (2014) The trajectories of adolescent anxiety and depressive symptoms over the course of a transdiagnostic treatment. Journal of Anxiety Disorders 28 (6):511-521. doi:10.1016/j.janxdis.2014.05.007
42. Lutz W, Leach C, Barkham M, Lucock M, Stiles WB, Evans C, Noble R, Iveson S (2005) Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. Journal of Consulting and Clinical Psychology 73 (5):904-913. doi:10.1037/0022-006x.73.5.904
43. Boswell JF, Kraus DR, Miller SD, Lambert MJ (2015) Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. Psychotherapy Research 25 (1):6-19. doi:10.1080/10503307.2013.817696

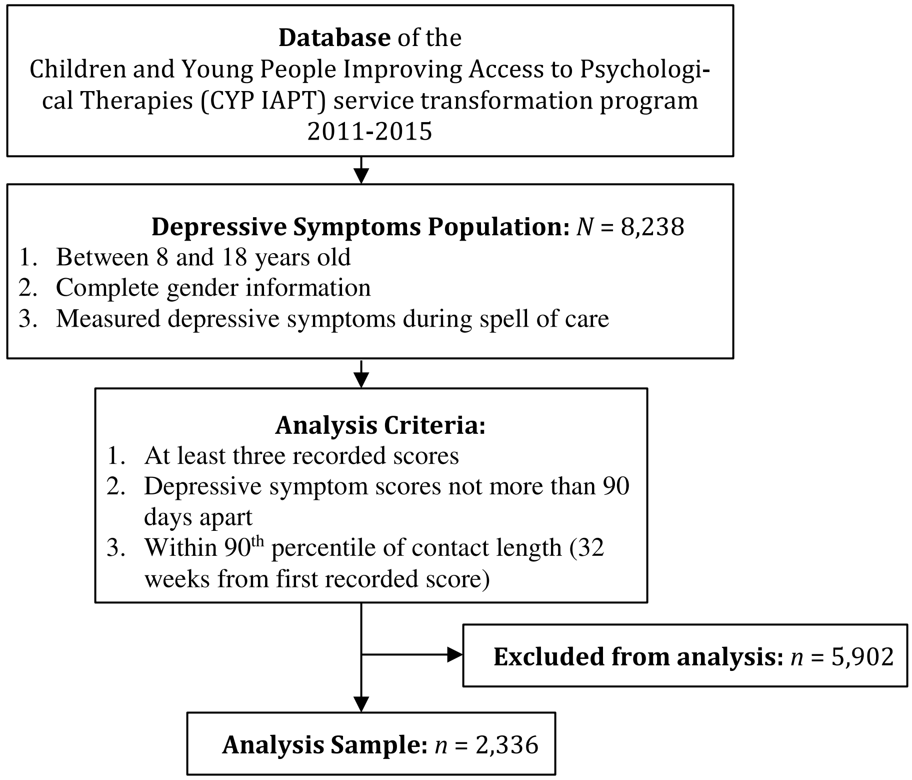

Figure 1. Sample selection flow

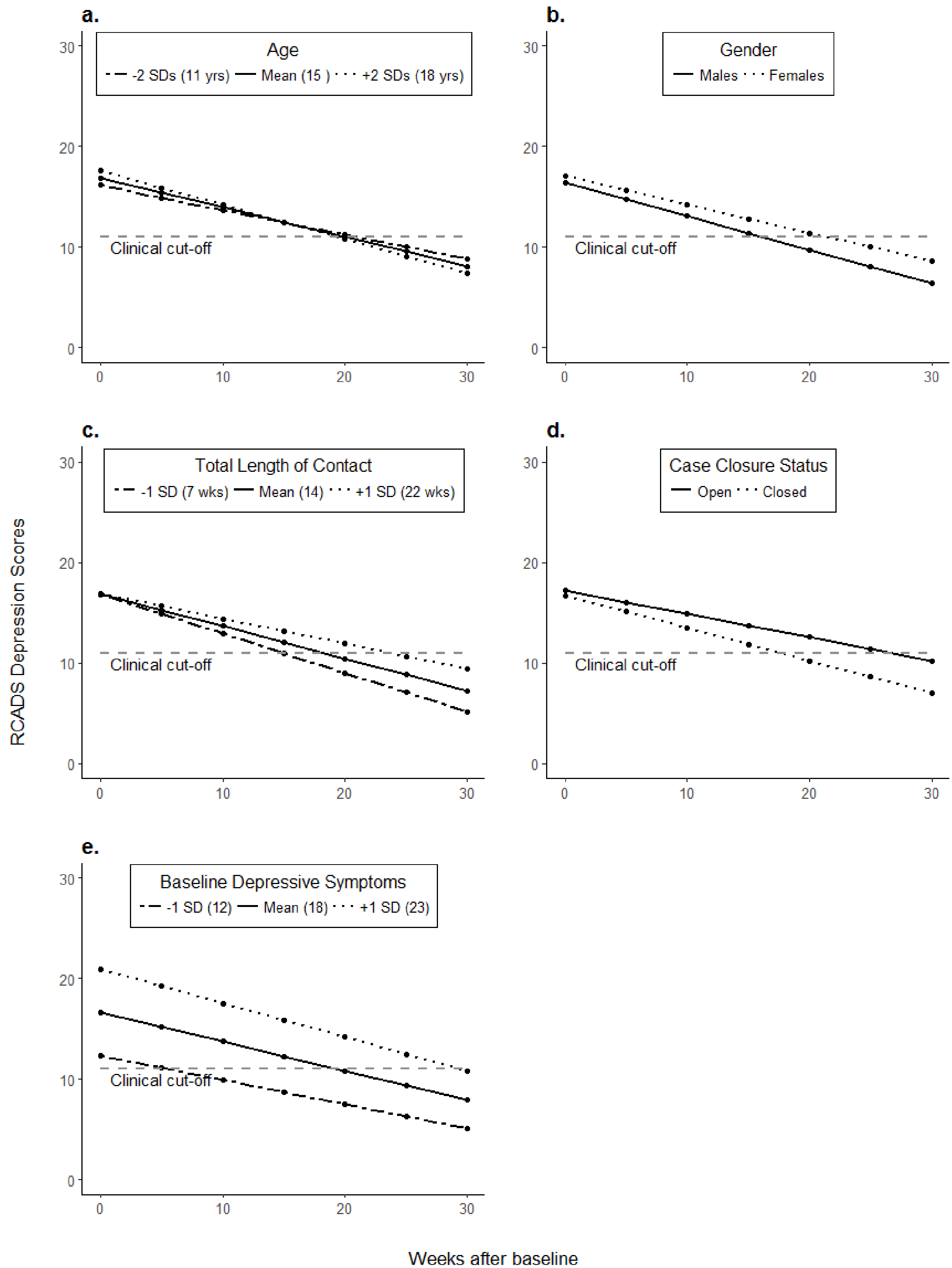

Figure 2. Estimated average trajectories of change in youth depressive symptoms

Table 1

Descriptive Characteristics of the Depressive Symptoms Sample and the Selected Sample

	Depressive Symptoms Sample			Selected Sample		
Variables	N	Mean / N	SD / %	N	Mean / N	SD / %
Age	8238	14	2.2	2336	14.5	1.7
Gender (Females)	8238	5757	69.9%	2336	1803	77.2%
Ethnicity (White)	5521	4513	81.7%	1549	1366	88.2%
Closed Cases	8238	6031	73.2%	2336	1536	65.8%
Length of Contact	8238	13.6	19.3	2336	14.4	7.2
Number of Sessions	8238	3.3	3.7	2336	5.7	3.1
Baseline Depression	8238	15.3	6.7	2336	17.7	5.8
% Above Cut-off at Base-line	8238	6155	74.7%	2336	2072	88.7%

Table 2

Steps Taken to Build the Change Trajectory Model

	Models	Definition	AIC	Deviance	*p*	
1	OLS Model	Ordinary Least Squares	72726	-36361	*NA*	
2	Model 1	Unconditional Means	65527	-32761	*NA*	
3	Model 2a	Unconditional Fixed Linear Slope Growth	64212	-32102	<0.001	***
4	Model 2b	Unconditional Random Linear Slope	62993	-31491	<0.001	***
5	Model 2c	Unconditional Fixed Log-linear Slope	64310	-32151	<0.001	***
6	Model 2d	Unconditional Random Log-linear Slope	63426	-31707	<0.001	***
7	Model 2e	Unconditional Fixed Square-root Slope	64237	-32115	<0.001	***
8	Model 2f	Unconditional Random Square-root Slope	63237	-31612	<0.001	***
9	Model 3	Random Linear Slope & Explanatory Variables (Demographics and Case Characteristics)	62762	-31365	<0.001	***
10	Model 4	Random Linear Slope & Explanatory Variables (Severity)	61210	-30587	<0.001	***
11	Final Model	Random Linear Slope & All Significant Explanatory Variables	61210	-30589	0.118	

*Note*s. *Models 2a, 2c and 2e* are nested within, and therefore compared to, *Model 1*; *Models 2b, 2d and 2f* are nested within the preceding one; *Model 3* is nested within *Model 2b*. All other models are nested within the preceding one. The Final Model excludes the explanatory variable of prior contact with the service, whose main effect and interaction were found to be non-significant.

*** $p < .001$

Table 3

Final Change Trajectory Model: Fixed and Random Effects

Fixed effects	Estimate	Std.Error	DF	T-value	*p*	
(Intercept)	16.48	0.23	8583	72.95	<0.001	***
Time in weeks	-0.3	0.03	8583	-10.12	<0.001	***
Age (Centered)	0.21	0.05	2330	4.08	<0.001	***
Female	0.6	0.22	2330	2.79	0.005	**
Length of Contact (Centered)	0	0.01	2330	0.39	0.694	
Closed Case	-0.53	0.19	2330	-2.8	0.005	**
Baseline Depression (Centered)	0.75	0.02	2330	46.3	<0.001	***
Time * Age (Centered)	-0.01	0.01	8583	-2.17	0.03	*
Time * Female	0.05	0.03	8583	2.01	0.044	*
Time * Total Length of Contact (Centered)	0.01	0	8583	5.71	<0.001	***
Time * Closed Case	-0.09	0.02	8583	-3.83	<0.001	***
Time * Baseline Depression (Centered)	-0.01	0	8583	-4.04	<0.001	***
Random effects	Estimate	SD	Corr.			
(Intercept)	13.41***	3.66	(Intr)			
Time in weeks	0.12***	0.35	0.006			
Within-subjects Residual	8.76***	2.96				

Notes. Continuous variables are centered around the grand mean. The Intercept value represents the average post-baseline score for cases that have 0 on all the variables included in the model: these are cases with a mean age of 15, male, with an average total length of treatment of 17 weeks, with a case status of "open", and with an average depression score of 18). Other estimates are to be interpreted as deviations from the Intercept value. The trajectories of interest are the interactions between Time and the explanatory variables. * p < .05; ** p < .01; *** p < .001