

## Short Report

# Assessing outcomes and satisfaction in a sustained national primary care youth mental health programme: a brief multi-methods report

Jeff Moore , Anna Blix , Jim Lyng and Joseph Duffy

Jigsaw – The National Centre for Youth Mental Health, Dublin, Ireland

### Abstract

**Objectives:** Most mental health difficulties have their onset in early adolescence. Increasingly, community based primary care is recognised as a critical pathway to early intervention. Despite encouraging initial evaluations, there is an ongoing need for evidence of the outcomes of primary care youth mental health programmes delivered at scale. This brief report examines reliable improvements in psychological distress and user satisfaction data from a national primary care youth mental health programme in the sustainment phase of implementation.

**Methods:** This report takes a multi-methods approach to routine evaluation data. Young people (aged 12–25;  $N = 8,721$ ) completed Clinical Outcomes Routine Evaluation (CORE-10 and YP-CORE) pre- and post-treatment. Clinical cut offs and a reliable change index (based on established guidelines) were used to report rates of reliable improvement. The analysis examined differences in outcomes based on age, gender, and clinical need. Satisfaction was measured using the youth service satisfaction survey ( $N = 4,267$ ). Natural language processing techniques were employed to objectively analyse qualitative user feedback.

**Results:** Most young people presented in the clinical range, with almost two-thirds reporting moderate to severe distress. Statistically significant reductions in distress were observed with large effect sizes ( $d = 1.08$ – $1.28$ ). Young people in the clinical range demonstrated significantly higher rates of reliable improvement compared to those who presented in the healthy range. In line with similar evaluations, young adults were more likely to achieve improvement and report higher satisfaction. Sentiment analysis of satisfaction data indicated a strong skew towards positive sentiment, with trust, anticipation and joy being predominant. Qualitative feedback pointed to waiting times as an improvement area.

**Conclusion:** The absence of a control group limits our ability to evaluate the effectiveness of the service interventions. Nonetheless after a decade of service delivery, these results indicate that large scale national youth mental health programmes can achieve satisfaction and clinical outcomes in line with international standards. Further research is needed on the predictors of reliable change, differences across demographic groups and approaches to improving waiting times in primary care.

**Keywords:** Primary care; reliable improvement; youth mental health

(Received 17 May 2024; revised 13 February 2025; accepted 14 March 2025)

### Introduction

About 50% of mental difficulties have their onset by the age of 14, and 75% by the age of 24 (Kessler *et al.* 2005; Uhlhaas *et al.* 2023). Primary care settings are increasingly recognised as pivotal entry points for young people in the early stages of mental health challenges. An international movement for community based mental health services for young people aged 12–25 years started around 20 years ago. Jigsaw (Ireland), Headspace (Australia) and Foundry (Canada) are examples of these services implemented at scale. Recent review studies of these initiatives show encouraging outcome data but also

emphasise the need for ongoing objective and peer-reviewed evaluations (Tuaf & Orkibi 2023; Settapani *et al.* 2017). Health service evaluation is also a critical component of Irish mental health policy (Department of Health 2020).

### Background

At early-stage implementation, Jigsaw: the National Centre for Youth Mental Health (Ire) reported a range of implementation and clinical outcomes (O’Keeffe *et al.* 2015). In this brief report, we set out to provide an objective assessment of improvements in mental health distress and user satisfaction with Jigsaw’s national primary care youth mental health programme at the sustainment phase of implementation. We also examine variations in outcomes by clinical need, over time and by age and gender to understand how different groups respond to this intervention and to inform future service improvements.

**Corresponding author:** Jeff Moore; Email: [jeff.moore@jigsaw.ie](mailto:jeff.moore@jigsaw.ie)

**Cite this article:** Moore J, Blix A, Lyng J, and Duffy J. Assessing outcomes and satisfaction in a sustained national primary care youth mental health programme: a brief multi-methods report. *Irish Journal of Psychological Medicine* <https://doi.org/10.1017/ipm.2025.18>

## Methods

### Participants

Participants were young people between the ages of 12–25 attending Jigsaw across a four year period. Young people generally received an initial screen, an assessment, and up to six sessions of evidence-informed mental health therapy (for full details of the clinical model, see O'Reilly *et al.* 2022). Only those who completed both a pre- and post-survey (and missing one item per scale or less) were included in the analysis ( $n = 8,721$ ). Additionally, 4,267 participants completed a satisfaction survey at the end of their treatment. All data collection was part of routine care with informed consent obtained prior to treatment. The study received approval from the Jigsaw Research Ethics Committee.

### Data collection

Mental health distress was measured via YP-CORE (Twigg *et al.* 2016; O'Reilly *et al.* 2016) for participants under 17, and the CORE-10 (Barkham *et al.* 2013; Connell & Barkham 2007) for those aged 17 and over. These 10-item measures demonstrated good internal consistency (CORE-10  $\alpha = 0.83$ , YP-CORE  $\alpha = 0.85$ ). Post-treatment participants were also asked to fill out the Youth Service Satisfaction Scale (YSSS;  $\alpha = 0.92$ ; Rickwood *et al.* 2019), which included questions on service and session satisfaction and two open-ended questions regarding what helped and areas for improvement.

### Data analysis

Similar to other evaluations of youth counselling services (see Duncan *et al.* 2020) we used the author-defined reliable change index and clinical cut off for the YP-CORE (Twigg *et al.* 2016) and CORE-10 (Barkham *et al.* 2013; Wise 2004) to benchmark outcomes against established standards. For the YP-CORE, changes were reported based on age and gender-specific criteria (Twigg *et al.* 2016; Blackshaw 2017). Differences in reliable change across these groups and over time were assessed using analysis of variance, chi squared and *t*-tests. Due to high volumes of missing data, analysis was not conducted on a number of protected characteristics (e.g. ethnicity).

To add objectivity to our analysis of qualitative user feedback, we applied natural language processing (NLP) techniques. Using the tidy text package in R studio, we cleaned and tokenised the data and removed stop words. Bigram analysis was conducted, followed by sentiment analysis using the Bing Liu sentiment lexicon to categorise responses as either positive or negative and the AFFIN tool to calculate sentiment intensity (Nielsen 2011).

## Results

### Clinical outcomes in routine evaluation

Of those who completed pre- and post-evaluation ( $n = 8,721$ ), 69.14% ( $n = 6,030$ ) were female and 29.15% ( $n = 2,543$ ) were male. The mean age of the sample was 15 years ( $SD = 3.01$ ). The majority were white Irish ( $n = 5,309$ , 60.87%). Most were referred by a parent or guardian (64.55%,  $n = 5,630$ ), followed by self-referral (18.93%,  $n = 1,651$ ), GP (6.44%,  $n = 562$ ), and 1% ( $n = 131$ ) were referred by Child and Adolescent Mental Health Services (CAMHS). The mean number of sessions for this group was 6.76 ( $SD = 2.36$ ). The mean wait time for an appointment for this sample was 10.05 weeks (range 0–43,  $SD = 6.72$ ). Of the full sample, 7,412 (85.23%) were in the clinical range at presentation.

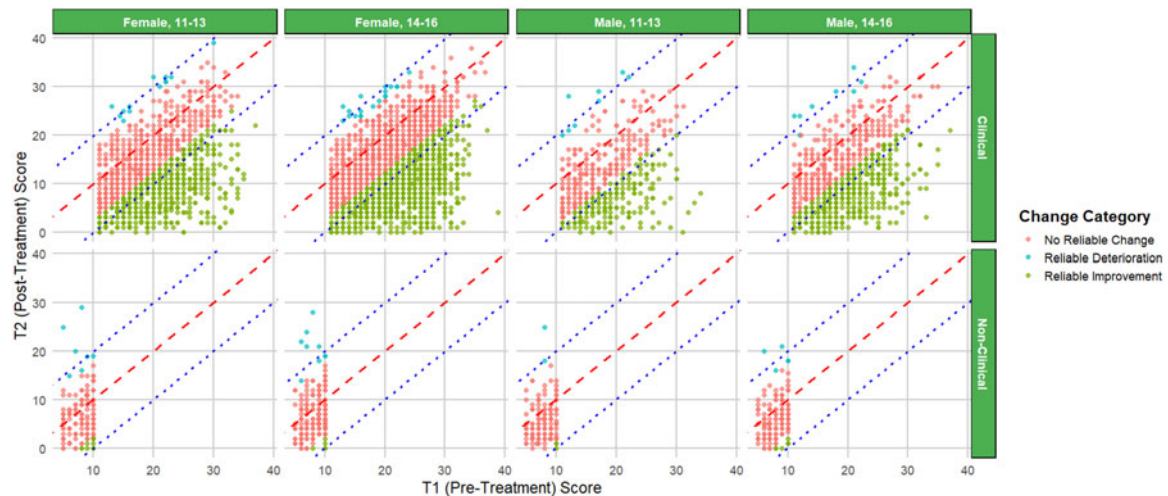
Of those, the majority (90.4%) completed the programme, with 4.05% recorded as partially completed programme and less than 1% (0.44%) recorded as inappropriate service based on needs. For the YP-CORE, the mean score at time one was 18.02 ( $SD = 7.57$ ). The CORE-10 mean at time one was 17.51 ( $SD = 6.51$ ). For the younger group, 85.96% ( $n = 8,570$ ) presented in the clinical range (moderate, moderate-severe and severe). For the older group (CORE-10) the majority of participants (2,679; 89.2%) presented in the clinical range.

A *t*-test showed a significant decrease in YP-CORE scores from time 1 to time 2,  $t(5721) = 75.82$ ,  $p = 0.001$ , with a mean difference of 6.88. The effect size, Cohen's  $d = 1.08$  (95% CI: [1.04, 1.11]), indicated a large effect. Figure 1 illustrates change categories by age and gender (as recommended by Twigg *et al.* 2016). A chi-square test ( $\chi^2 = 530.06$ ,  $p < 0.001$ ) showed a significant difference in reliable improvement between clinical and non-clinical groups. For clinical participants ( $n = 4,785$ ), 51.5% ( $n = 2,465$ ) showed reliable improvement, 47.5% ( $n = 2,272$ ) showed no reliable change, and 1% ( $n = 48$ ) experienced reliable deterioration. For non-clinical participants ( $n = 672$ ), 92% ( $n = 618$ ) showed no reliable change. Females showed higher rates of reliable improvement across age groups, with 43.8% of females aged 11–13 ( $n = 590$ ) and 49.7% of females aged 14–16 ( $n = 1,267$ ) showing improvement, compared to 33.9% of males aged 11–13 ( $n = 204$ ) and 45.3% of males aged 14–16 ( $n = 435$ ). A chi-square test revealed significant differences in change categories between males and females,  $\chi^2(n = 5,457) = 20.87$ ,  $p < .001$ . ANOVA results show a statistically significant difference in scores across years  $F(3, 5718) = 5.483$ ,  $p < .001$ . Post hoc analysis showed a significant difference between 2021 and 2022 with an estimate of 0.93, suggesting that on average, the reliable improvements scores in 2022 were higher than in 2021.

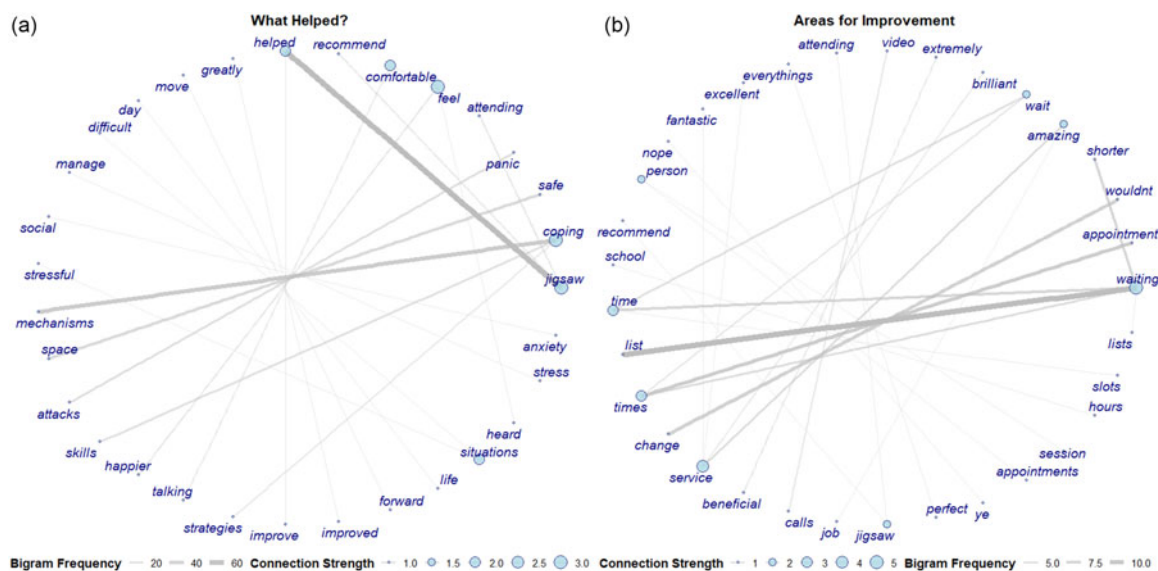
A *t*-test demonstrated a significant reduction in distress between time 1 and time 2 on the CORE-10,  $t(3003) = 65.286$ ,  $p < 0.001$ , with a mean difference of 7.49 and large effect size ( $d = 1.24$ ). A Pearson's chi-square test again showed a significant difference between clinical ( $n = 2,679$ ) and non-clinical ( $n = 325$ ) groups ( $\chi^2 = 368.33$ ,  $df = 2$ ,  $p < 0.001$ ), with 62.37% ( $n = 1,671$ ) of clinical participants improving compared to 6.46% ( $n = 21$ ) in the non-clinical group, and 36.28% ( $n = 972$ ) of clinical versus 90.15% ( $n = 293$ ) of non-clinical participants showing no reliable change. As with the YP-CORE, less than five percent showed a deterioration. ANOVA results showed a significant difference in rates of reliable improvement by age ( $F(1, 9062) = 18.95$ ,  $p < 0.001$ ), with young people in the 19–25 age group achieving significantly higher rates of reliable improvement (mean = 7.83) compared to those in the 17–18 age group (mean = 6.99). An ANOVA showed a significant difference in reliable improvement scores between genders,  $F(1, 8906) = 16.23$ ,  $p < 0.001$  with females ( $M = 7.32$ ,  $SD = 6.83$ ) scoring a significantly higher mean change score compared to males ( $M = 6.69$ ,  $SD = 6.54$ ). An ANOVA examined the differences in reliable improvement by calendar year,  $F(3, 9060) = 6.116$ ,  $p = .003$ . Post hoc comparisons showed a significant decrease between 2021 and 2022 ( $p < .001$ ), while a significant increase was found between 2022 and 2023 ( $p = 0.02$ ).

### User satisfaction and sentiment analysis

Females made up two thirds (66%) of those who completed the satisfaction survey and the average age of respondents was 17 years. Almost ninety percent (87.20%) agreed or strongly agreed that



**Figure 1.** Categories of distress pre- and post-interventions (2020–2023). YP-CORE clinical outcomes stratified by age, gender, and clinical status.



**Figure 2.** Common Bigrams in youth feedback (a) what helped (b) what could be improved?.

their mental health had improved, and 82.89% agreed or strongly agreed that their lives had improved. The overall mean for the YSSS was 4.45 ( $SD = 0.49$ ). Other studies have reported group means ranging from 3.05 to 4.56 (Doyle *et al.* 2024, Rickwood *et al.* 2019). A paired samples *t*-test showed no significant differences in scores between females and males,  $t(2494) = 0.09$ ,  $p = 0.923$ . In comparing scores between the younger (12–16) and older (17–25) age groups, we found the older age group showed significantly higher levels of satisfaction:  $t(4262) = -10.27$ ,  $p < 0.001$ . An ANOVA revealed a significant difference in YSSS scores across years ( $F(3, 4260) = 6.099$ ,  $p < .001$ ). The post-hoc tests indicate that YSSS scores significantly decreased between 2020 and 2022 ( $p < .001$ ) as well as between 2020 and 2023 ( $p < .001$ ). No significant difference was observed between 2020 and 2021 ( $p = 0.471$ ). Figure 2 illustrates the most common bigrams in the qualitative data. Bigram frequency represents how

often two words co-occur, while connection strength indicates the importance of a word in linking others within the network. In responding to what helped, ‘jigsaw helped’ and ‘coping mechanisms’ were the most common bigrams, followed by ‘coping skills’. Alongside coping mechanisms, bigrams pointed to positive outcomes (e.g. ‘feel happier’). Bigrams like ‘safe space’ and ‘comfortable talking’ highlights the importance of environment in improving satisfaction. In terms of areas for improvement, ‘waiting’ was the primary issue, with concerns about waiting times, appointments, and scheduling, supported by secondary terms like ‘list’, ‘shorter’ and ‘times’.

Using the AFFIN sentiment analysis, 60.27% of words were identified as positive, while 39.73% ( $n = 4,188$ ) were classified as negative ( $n = 2,761$ ). A sentiment intensity score of 0.654 indicated a strong skew towards positive sentiment. Emotional intensity was analysed via the NRC package (ranging from 0–1). This showed



'trust' had the highest average intensity score at 0.538, 'anticipation' followed closely at 0.501 and 'Joy' (0.477), suggesting a positive sentiment overall. Other emotions like 'anger', 'fear', and 'sadness' show moderate levels of expression with scores of 0.370, 0.382, and 0.437.

## Discussion

Overall, our results indicate that the Jigsaw service is associated with mental health improvements for the majority of users. Our findings indicate that the rates of reliable improvements in young people compare very well with previous research in similar primary care settings and using a similarly strict RCI (37.2% in Duncan *et al.* 2020; 55.9% in Twigg *et al.* 2016). A range of studies report reliable improvements in children and young people in IAPT services at approximately 50% (Edbrooke-Childs *et al.* 2018; Gyani *et al.* 2013; Wolpert *et al.* 2016; NHS Digital 2022). We found very similar rates of reliable improvements for the older group (17–25) to those reported by adult IAPT services (64%; Gyani *et al.* 2013). Our effect size for change compares favourably with those reported in other large-scale primary care talk therapy evaluations (Brand *et al.* 2021).

Results demonstrate high levels of satisfaction irrespective of gender. Analysis of qualitative feedback revealed trust and joy as dominant emotions. User feedback indicated that participants valued the support they received, particularly the safe, comfortable environment and the development of coping mechanisms, while improvements are needed in reducing waiting times and enhancing appointment scheduling. Similar to recent evaluations of Headspace services (Headspace National Youth Mental Health Foundation 2022), our results show that young adults were more satisfied and more likely to achieve reliable improvements.

While the size of the sample is a strength of the study, it is limited by the lack of a control group and this limits our ability to comment on the effectiveness. The use of a standardised RCI, while a strength in terms of comparisons, reduces the precision of our analysis and may result in under reporting of improvements. Moreover, while the CORE provides a key indicator of clinical progress it does not capture more holistic outcomes, such as social functioning. An examination of outcomes by presenting complexity or demographics was not possible due to high levels of missing data which is a serious limitation of this study. NLP, while objective, is limited in terms of the richness and depth of analysis and further analysis of these data is warranted.

## Conclusion

This brief report illustrates that as Jigsaw reaches sustainment phase, the programme continues to achieve high level of user satisfaction and rates of reliable improvement beyond or in line with other national programmes. Differences across clinical need, age and gender highlight the importances of personalised approaches to care and also signal the need for further research examining access and outcomes across groups and improving waiting times.

**Financial support.** The work of Jigsaw's research and evaluation department is funded by the health service executive (HSE, Ireland).

**Competing interests.** At time of writing, all authors were employed by Jigsaw; the National Centre for Youth Mental Health. Authors declare no other conflicts of interest.

This manuscript is not under consideration with any other publication.

**Ethical standards.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## References

- Barkham M, Bewick B, Mullin T, Gilbody S, Connell J, Cahill J, Mellor-Clark J, Richards D, Unsworth G, Evans C (2013). The CORE-10: a short measure of psychological distress for routine use in the psychological therapies. *Counselling and Psychotherapy Research* 13, 3–13.
- Blackshaw E (2017). *When life gets in the way: A systematic review of life events, socioeconomic deprivation, and their impact on counselling and psychotherapy with children and adolescents (Doctoral dissertation)*. University of Roehampton.
- Brand C, Ward F, MacDonagh N, Cunningham S, Timulak L (2021). A national evaluation of the Irish public health counselling in primary care service– examination of initial effectiveness data. *BMC Psychiatry* 21, 227. doi:10.1186/s12888-021-03226-x.
- Connell J, Barkham M (2007). *CORE-10 User Manual, Version 1.1*. CORES system Trust & CORE Information Management Systems Ltd
- Department of Health (2020). *Sharing the vision: A mental health policy for everyone*. Government of Ireland: Dublin.
- Doyle E, Carey E, Rossouw J, Booth A, Rickwood D, O'Reilly A (2024). Examining the psychometric properties of the headspace Youth (mental health) Service Satisfaction Scale in a mental health service in Ireland. *Child Care Health and Development*, 50.
- Duncan C, Rayment B, Kenrick J, Cooper M (2020). Counselling for young people and young adults in the voluntary and community sector: An overview of the demographic profile of clients and outcomes. *Psychology and Psychotherapy: Theory, Research and Practice*, 93(1), 36–53. Doi: 10.1111/papt.12206.
- Edbrooke-Childs J, 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, 250–255. doi:10.1192/bjo.2018.31.
- Gyani A, Shafraan R, Layard R, Clark DM (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy* 51, 597–606. doi:10.1016/j.brat.2013.06.004.
- Headspace National Youth Mental Health Foundation (2022). Young people snapshot report. Retrieved from [https://headspace.org.au/assets/HSTC184\\_-Young-People-Snapshot-Report\\_-FA.pdf](https://headspace.org.au/assets/HSTC184_-Young-People-Snapshot-Report_-FA.pdf)
- Kessler RC, Chiu WT, Demler O, Walters EE (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62, 617–627.
- NHS Digital (2022). Improving Access to Psychological Therapies (IAPT) Key Performance Indicator (KPI) Report. NHS Digital. <https://digital.nhs.uk/data-and-information/publications/statistical/improving-access-to-psychological-therapies-report>
- Nielsen FÅ (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC Workshop on 'Making Sense of Microposts'*. CEUR Workshop Proceedings, vol. 718. [http://ceur-ws.org/Vol-718/paper\\_16.pdf](http://ceur-ws.org/Vol-718/paper_16.pdf)
- O'Keeffe L, O'Reilly A, O'Brien G, Buckley R, Illback R (2015). Description and outcome evaluation of Jigsaw: an emergent Irish mental health early intervention programme for young people. *Irish Journal of Psychological Medicine*, 32(1), 71–77. doi: 10.1017/ipm.2014.86.
- O'Reilly A, O'Brien G, Moore J, Duffy J, Longmore P, Cullinan S, Mc Grory S (2022). Evolution of Jigsaw – a National Youth Mental Health Service. *Early Intervention in Psychiatry* 16, 561–567.

- O'Reilly A, Peiper N, O'Keeffe L, Illback R, Clayton R (2016). Performance of the CORE-10 and YP-CORE measures in a sample of youth engaging with a community mental health service. *International Journal of Methods in Psychiatric Research* 25, 324–332.
- Rickwood D, Wallace A, Kennedy V, O'Sullivan S, Telford N, Leicester S (2019). Young people's satisfaction with the online mental health service eheadspace: Development and implementation of a service satisfaction measure. *JMIR Mental Health* 6(4), e12169. doi:[10.2196/12169D](https://doi.org/10.2196/12169D).
- Settipani CA, Cleverley K, Hawke LD, Rice M, Henderson JL (2017). Essential components of integrated care for youth with mental health and addiction needs: protocol for a scoping review. *BMJ Open* 7, e015454. doi:[10.1136/bmjopen-2016-015454](https://doi.org/10.1136/bmjopen-2016-015454).
- Tuaf H, Orkibi H (2023). Community-based programs for youth with mental health conditions: a scoping review and practical implications. *Frontiers in Public Health* 11, 1241469.
- Twigg E, Cooper M, Evans C, Freire E, Mellor-Clark J, McInnes B, Barkham M (2016). Acceptability, reliability, referential distributions and sensitivity to change in the Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE) outcome measure: replication and refinement. *Child and Adolescent Mental Health* 21, 115–123. doi:[10.1111/camh.12128](https://doi.org/10.1111/camh.12128).
- Uhlhaas PJ, Davey CG, Mehta UM, Shah J, Torous J, Allen NB *et al.* (2023). Towards a youth mental health paradigm: a perspective and roadmap. *Molecular psychiatry* 28, 3171–3181.
- Wise EA (2004). Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of personality assessment* 82, 50–59. doi:[10.1207/s15327752jpa8201\\_10](https://doi.org/10.1207/s15327752jpa8201_10).
- 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.