Evaluation of reliable improvement rates in depression and anxiety at the end of treatment in adolescents

Julian Edbrooke-Childs*, Miranda Wolpert*, Victoria Zamperoni, Elisa Napoleone and Holly Bear

Background

Literature has focused on effect sizes rather than individual-level improvement rates to determine how effectively services address burgeoning numbers of adolescents with anxiety and depression.

Aims

To consider how many adolescents report reliable improvement in anxiety, depression and comorbid depression and anxiety by end of treatment.

Method

The primary outcome was reliable improvement (i.e. change greater than likely the result of measurement error) in self-reported anxiety and depression for N = 4464 adolescents (mean age 14.5 years, s.d. = 1.9; 75% female; 61% White) seen in specialist mental health services in England.

Results

In total, 53% of those with anxiety, 44% with depression, and 35% with comorbid depression and anxiety showed reliable improvement.

Conclusions

Improvement rates were higher than previously reported, but lower than generally used in advice to the public. There may be a need to set more realistic expectations, including with young people who seek help.

There is evidence of increasing levels of anxiety and depression in adolescents internationally,[1–7] leading to an emphasis on the need for more treatment access. However, there is much less discussion or evidence about how many are substantially better following treatment nor how to discuss the limitations of treatment outcomes. The dominant message currently is to encourage children to seek treatment and to emphasise the likely positive outcomes from treatment. The most respected current guidance offers no reference to likely improvement rates for those accessing therapy for anxiety and/or depression. For example, in the UK, in the National Institute for Health and Care Excellence public recommendations on treatment for anxiety and depression in children and young people, there is no mention of improvement rates and no quality standards in relation to improvement expected to be achieved following access to services.[8,9] Nor is any mention of this made in guidance from the Royal College of Psychiatrists.[10] Websites offering information to the public stress the need for help and suggest that without help children and adolescents will not improve, but with help they will. For example, according to the website of the Anxiety and Depression Association of America: 'Like other medical conditions, anxiety disorders tend to be chronic unless properly treated. Most kids find that they need professional guidance to successfully manage and overcome their anxiety'.[11]

Outcomes of treatment are generally reported in terms of effect size differences between treatments[12] or between before and after treatment, and not in terms of individual-level improvement rates. Recently, use of effect sizes has been critiqued as potentially contributing to overinflated estimates of impact for a given individual as it does not include any consideration of clinically meaningful metrics of change, nor does it take into account differences between individuals in terms of outcomes, and that focusing on mean differences can be misleading in terms of the likely impact on individuals. This may be true in a number of ways; for example, if a small number of people improve a lot and a large number of people improve a little (or even some get worse), the mean result may look positive. Alternatively, everyone could improve but by such small amounts as to be not meaningful in relation to their daily lives. Thus, looking at effect size as a metric might not be the most useful statistic and, in particular, may not be of help to people providing or receiving treatment.[13] There has therefore been a call for a focus on metrics of outcome that focus on individual responses to treatment that may be more clinically relevant and meaningful.[14]

The metric of 'reliable change' is increasingly being suggested for use[15–18] as an estimate of potentially clinically important change at the individual level. Reliable change refers to the amount of change in scores on a scale; it 'tells us whether change reflects more than the fluctuations of an imprecise measuring instrument'.[19] This metric considers whether improvement (or indeed deterioration) in symptoms is greater than could likely be solely attributed to measurement error.

The systematic review of child mental health research conducted for this article found only five studies that examined self-reported reliable improvement following exposure to routine care

---

* These authors are joint first authors.

for young people with anxiety and/or depression. These have all been undertaken in the USA, Australia or Scandinavia and employed small samples (n = 11–85 adolescents) and report low rates of reliable improvement (8–36%) with 0–10% deterioration (see Table 4).

There is thus an important gap in the literature about individual-level rates of improvement following care. This is crucial information to help practitioners set realistic expectations for outcomes and support adolescents to make informed choices about their care. The aim of the present research was to address this gap by considering how many adolescents' reports indicate reliable improvement in anxiety, depression and comorbid depression and anxiety by end of treatment.

Method

Participants and procedure

Data were collected from 75 services where practitioners were trained in key evidence-based interventions and use of feedback and outcome monitoring (Children and Young People's Improving Access to Psychological Therapies (2011–2015)[20]).

Services were part of publicly funded provision and staffed by multidisciplinary teams including nurses, psychologists, psychiatrists, counsellors and others. The programme emphasised provision of evidence-based interventions for given problems such as cognitive–behavioural therapy for anxiety and depression or interpersonal therapy for depression but the majority of treatment provided drew on a range of interventions.

The treatment episodes' participants were deemed eligible for analysis if they were recorded as 'closed' and had at least three recorded events (at least one of which was for a purpose other than assessment), n = 27 439; the participants had completed a self-reported measure of anxiety or depression (see Measures) at assessment, n = 13 885; was between the age of 8 and 18, n = 13 599; and scored above the threshold at assessment for their school year and gender on either the anxiety (i.e. social phobia, separation anxiety, panic disorder, or generalised anxiety) or depression subscales, n = 9028 episodes of care. Of these, 8927 were unique individuals and 101 were re-referrals of the same individual in this period.

Note that it was not possible to be certain that the last available record was at the same time point as end of treatment but we can be sure it was the last recorded measure by end of treatment and that it was further to the first recorded measure.

Data were then filtered to include only those episodes of treatment where the same measure of anxiety or depression completed at assessment was also completed on more than one occasion in the same episode of treatment (referred to as 'paired' data). Scores from the first-ever and last-ever recorded completion of the measures were then analysed to assess change in self-reported anxiety and depression symptoms over time. This resulted in a final included sample of n = 4464 episodes of care (49.4% of the time 1 total sample, from 70 services).

Demographic characteristics for the total and final samples are presented in Table 1. There were some significant differences in age, gender and ethnicity between those episodes of treatment from the final sample and the total sample but no differences in Revised Child Anxiety and Depression Scale (RCADS)[21] scores. All differences found displayed small effect sizes. Young people in the final sample were slightly older compared with those in the total sample (odds ratio (OR) = 1.2, 95% CI 1.1–1.3), slightly more likely to be female (OR = 1.2, 95% CI 1.1–1.3), slightly less likely to be of mixed ethnicity (OR = 0.7, 95% CI 0.5–0.9), and slightly more likely to have not stated their ethnicity (OR = 1.2, 95% CI 1.1–1.4). As measures were taken from a secondary analysis of anonymised routinely collected data, ethical review was not required.[22]

Measures
Demographic characteristics

Age, gender and ethnicity were recorded by services as part of routine data recording. Ethnicity was captured based on self-report by the carer or adolescent and were grouped as follows: White (including White British, Irish, and other White background), mixed (including mixed White and Black Caribbean, mixed White and Black African, mixed White and Asian, and any other mixed background), Asian or Asian British (including Indian, Pakistani, Bangladeshi and other), Black or Black British (including Caribbean, African and other) and other ethnic groups (including Chinese and other).

Table 1	Descriptive characteristics of episodes of care with data at assessment only compared with those with 'paired' data					
	Data at assessment			Paired data		Odds ratio[a] (95% CI)
	n	Mean (s.d.)	%	n	Mean (s.d.)	%	
Age, years	4564	14.3 (2.2)		4464	14.5 (1.9)		1.2 (1.1–1.3)
Gender							
Male	1300		28.5	1120		25.1	1.2 (1.1–1.3)
Female	3264		71.5	3344		74.9	
Ethnicity							
White	2885		63.2	2721		61.0	0.9 (0.8–1.0)
Black	132		2.9	97		2.2	0.8 (0.6–1.0)
Mixed	163		3.6	111		2.5	0.7 (0.5–0.9)
Asian	96		2.1	87		1.9	1.0 (0.7–1.2)
Other	75		1.6	81		1.8	1.1 (0.8–1.5)
Not stated	883		19.3	1008		22.6	1.2 (1.1–1.4)
Missing	330		7.2	359		8.0	1.1 (1.0–1.3)
Baseline RCADS t-scores							
Separation anxiety	4389	70.4 (17.1)		4177	71.3 (16.3)		1.1 (1.0–1.2)
Social phobia	4380	60.2 (12.1)		4187	61.4 (11.8)		1.2 (1.1–1.3)
Panic	4362	72.0 (17.5)		4172	73.6 (17.5)		1.2 (1.1–1.3)
Generalised anxiety	4423	59.9 (11.6)		4186	60.7 (11.1)		1.1 (1.0–1.2)
Depression	4432	74.5 (14.3)		4296	75.0 (14.3)		1.1 (1.0–1.1)

RCADS, Revised Child Anxiety and Depression Scale.
a. Log odds ratios for continuous variables were calculated using Cohen's d based on independent sample t-tests, using the formula log(o) = (pi d)/(sqrt(3)).

Anxiety and depression

To measure the presence and severity of anxiety and depressive symptoms, the 47-item RCADS[21] was used. The RCADS measures symptoms of anxiety and depression and is composed of six subscales (depression, social phobia, separation anxiety, generalised anxiety, panic disorder and obsessive–compulsive disorder). Items are rated on a four-point scale from 0 (never) to 3 (always). In addition to the full 47-item questionnaire, the individual subscales were also used as shorter, stand-alone measures. The subscale for obsessive–compulsive disorder was excluded from this analysis as the literature suggests that the aetiology and course of this disorder may be different from the other four anxiety disorders measured by the RCADS.[23]

The RCADS is a widely used measure of child anxiety and depression, with evidence of validity and reliability (Cronbach's α = 0.78–0.88 across the six subscales).[21] According to data collection protocol, the full RCADS should be completed at initial assessment, then again 4–6 months later, or at case closure if sooner. The subscales of the RCADS may be used as stand-alone measures more frequently, and at the discretion of the practitioner, in order to track progress on specific symptoms throughout the episode of care.[24] The RCADS was found to be sensitive enough to pick up a change in anxiety and depression symptoms experienced by children and adolescents up to 12 months after receiving a preventative intervention.[25] Thresholds were based on norms from a study of school children in the USA.[22] In line with recommendations in the RCADS user guide, children were considered above the threshold if they had a t-score of 65 or greater for children of their school year and gender, indicating a score in the top 6% of the sample.[26]

Analytic strategy

First, we examined the primary outcome of whether change in scores was greater than could likely be solely attributed to measurement error by calculating the reliable change criterion.[19] A reliable change criterion was calculated for each subscale of the RCADS as shown in the formulae below, using the internal consistency of the subscale and the s.d. of the final included sample at the first time point.

$$\text{Reliable change criterion} = \text{SE}_{\text{diff}} \times 1.96, \text{ where}$$

$$\text{SE}_{\text{diff}} = \text{s.d.} \times \sqrt{2} \times \sqrt{1 - \text{reliability}}$$

For those scoring above the threshold at assessment, participants that exhibited improvement in scores greater than the reliable change criterion were deemed 'reliably improved', participants with deterioration in scores greater than the reliable change criterion were deemed 'reliably deteriorated', and individuals not crossing the reliable change criterion were categorised as showing 'no reliable change'. To avoid multiple classification of participants when analysing anxiety individually and when analysing comorbid depression – where one patient could have both reliably improved and reliably deteriorated (or not reliably changed) on different subscales – the individual was considered reliably improved if they showed reliable improvement on at least one elevated subscale and no reliable deterioration on any of the other subscales. Conversely, a patient's condition was classified as reliably deteriorated if they had reliably deteriorated on any subscale. For participants with comorbid depression and anxiety, both elevated scales had to show reliable improvement for the individual to be classified as reliably improved.

Secondary analysis considered whether scores crossed to below cut-off at follow-up (i.e. 'recovered') or remained above the cut-off at follow-up (i.e. 'non-recovered'). To avoid multiple classification

of participants when analysing anxiety individually and when analysing comorbid depression and anxiety – where one patient could be both recovered and non-recovered on different subscales – a patient needed to be above the threshold on one subscale at assessment but below the threshold on all subscales at follow-up for their condition to be classified as recovered.

A further secondary analysis used the most conservative approach to examining change by assessing 'reliable recovery'. Here, participants that crossed from above the thresholds at assessment to below at follow-up and which demonstrated an improvement in score greater than the reliable change threshold on the same subscale (i.e. reliably improved) were classified as 'reliably recovered'.

We performed logistic regressions to examine whether demographic characteristics were associated with reliable improvement, recovery and reliable recovery. Most demographic characteristics did not significantly predict treatment outcome; however, gender was a significant predictor of recovery for the anxiety and depression comorbid group, although the regression model was only able to explain 1% of the variance in recovery (see Supplementary Tables 1–3, available at https://doi.org/10.1192/bjo.2018.31).

In addition, pre–post effect size calculations were carried out for each subscale based on the sample of participants who were above the threshold for each scale. It should therefore be noted that the sample used to calculate these effect sizes were not the same as the samples used for the main analyses. These effect size calculations were included for the sole purpose of aiding discussion of how individual-level metrics might be considered alongside results from effect size metrics. Effect sizes were calculated as ($\text{Mean}_{\text{Time 1}}$ – $\text{Mean}_{\text{Time 2}}$)/$\text{s.d.}_{\text{Time 1}}$. The CIs were estimated with the R package 'boot'[27] using 10 000 bootstrap samples.

Results

The results of the analyses are shown in Tables 2 and 3.

Anxiety

From the final sample of $n = 4464$ participants, 1208 were above the threshold on at least one of the anxiety subscales at assessment and not above the threshold for depression. The reliable change indices in t-scores were 12.26 for generalised anxiety, 16.78 for panic disorder, 21.17 for separation anxiety, and 11.75 for social phobia.

Table 2 Reliable change, recovery and reliable recovery in anxiety, depression and comorbid samples

	Anxiety $n = 1208$		Depression $n = 621$		Comorbid depression and anxiety $n = 2635$	
	n	%	n	%	n	%
Reliable change						
Reliable improvement	638	52.8	275	44.3	913	34.6
No reliable change	400	33.1	323	52.0	751	28.5
Reliable deterioration	93	7.7	23	3.7	256	9.7
Recovery						
Recovered	736	60.9	350	56.4	1020	38.7
Non-recovered	472	39.1	271	43.6	1615	61.3
Reliable recovery						
Recovered	555	45.9	258	41.5	672	25.5
Non-recovered	653	54.1	363	58.5	1963	74.5

a. Participants were considered to be above the threshold if they had a t-score of 65 or greater for their school year and gender, based on developers' norms.

Table 3 Pre–post effect sizes and 95% CIs of the Revised Child Anxiety and Depression Scale (RCADS) subscales[a]

RCADS subscales	*n*	Pre–post effect size	(95% CI)
Social phobia	1500	1.0	(1.0–1.1)
Separation anxiety	2081	0.8	(0.8–0.9)
Panic	2229	0.9	(0.9–1.0)
Generalised anxiety	1149	1.4	(1.3–1.4)
Depression	3052	0.9	(0.9–0.9)

a. The sample sizes reflect the number of closed episodes of care with a score above the subscale threshold at a first time point.

Depression

From the final sample of *n* = 4464 participants, 621 were above the threshold for depression and not above the threshold for any anxiety problems at assessment. The reliable change criterion for depression was 14.29, meaning individuals with *t*-scores that changed by more than 14.29 points were classified as having reliably changed.

Comorbid depression and anxiety

From the final sample of *n* = 4464 participants, 2635 were above the threshold for the depression subscale and at least one anxiety subscale at assessment. Of participants above cut-off for the depression subscale and at least one anxiety subscale at assessment, 913 (i.e. 34.6%): (a) reliably improved on the depression subscale, (b) reliably improved on at least one anxiety subscale that was above the threshold at assessment, and (c) had not reliably deteriorated on any subscale.

Pre–post effect sizes

Pre–post effect sizes for each subscale were calculated, drawing in each instance on the sample of children who were above the threshold at outset for that subscale. These are shown in Table 3.

As can be seen, pre–post effect sizes ranged from 0.8 (for separation anxiety, 95% CI 0.8–0.9) to 1.4 (for generalised anxiety, 95% CI 1.3–1.4). These would be deemed 'large' effect sizes.[28]

Discussion

Main findings

In the present research, 53% of young people with anxiety and no comorbid depression reliably improved, 44% with depression and no comorbid anxiety reliably improved and 35% with comorbid depression and anxiety reliably improved. The difference between anxiety, depression and comorbid depression and anxiety were in line with earlier findings, although the overall rates of reliable improvement were somewhat higher than in the five previous studies identified in our systematic review (see Table 4).

Interpretation of our findings

These rates of individual improvement are in contrast with how the results might have been conceived if this paper had focused on pre–post effect size analyses. These effect sizes, which ranged from 0.8 to 1.4, could have been deemed large effect sizes drawing on rules of thumb as described by Cohen.[28] Although many (including the current authors) would argue that such a designation would be inappropriate since these rules of thumb do not apply to pre–post measurements because of the non-independence of the data, and the fact that the s.d. chosen to calculate the effect size often results in changes of over 0.4 in the effect size itself,[34] in practice these rules are extensively applied also for pre–post effect sizes (for example Edlund *et al*[29]). The lack of rules of thumb for meaningful use of effect sizes pre–post adds to the challenges of using such statistics where no benchmarks exist.

It is important to note that these outcomes only appear modest when considered relative to the expectations that are being engendered by public-facing discourse and a lack of professional discussion of the fact that child mental health treatment will not help everyone (as noted above). In fact, these sort of rates of reliable improvement are closely in line with reliable improvement rates found in other psychological therapy for adults[15] and are higher than the modest extant literature of comparable studies on psychological therapy for adolescents and higher than many physical health outcomes for conditions that may be seen as chronic, such as paediatric diabetes where positive recorded outcome rates (based on control of haemoglobin A_{1c} to under a threshold of 7.5) have risen from around 15% in 2009–10 to around 24% in 2014–15.[35] One of the implications of this approach is that it might form the basis for more considered discussions around realistic expectations for treatment.

Limitations

There are clearly limitations to this research.[20,36] Only 49% of the adolescents with self-report measures at assessment had data at case closure so this analysis is based on a sample with much missing data. It is highly likely that there will be skews in the data in terms of who completed the questionnaires at both time points even though it should be noted that there were not differences found in the level of symptoms on RCADS scores between samples at outset, suggesting that this at least was not a source of skewing the data. We know from the sample analysis that adolescents who completed self-reported measure at both time points were generally somewhat older and less ethnically diverse than those who completed any child- or parent-reported assessment.[20] The question is whether this might have substantially affected the improvement rates. For example, it could be that those with the most improvement ended therapy early and did not complete a measure at a second time point, thus decreasing the overall improvement rates. Alternatively, it could be that those who achieved most in therapy were most motivated to complete the

Table 4 Reliable improvement and deterioration in anxiety, depression and comorbid anxiety and depression from the present research and previous studies

	Anxiety, % (*n/N*)				Depression, % (*n/N*)				Comorbid anxiety and depression or mixed emotional problems, % (*n/N*)	
	This study	Edlund *et al* (2016)[29]	Jónsson *et al* (2015)[30]	Biegel *et al* (2009)[31]	This study	Hayes *et al* (2011)[32]	Edlund *et al* (2016)[29]	Biegel *et al* (2009)[31]	This study	Nilsen *et al* (2015)[33]
Reliable improvement	53 (638/1208)	36 (19/53)	28 (24/85)	20 (8/40)	44 (275/621)	36 (4/11)	27 (14/51)	28 (11/40)	35 (913/2635)	8 (1/13)
Reliable deterioration	8 (93/1208)	8 (4/53)	6 (5/85)	3 (1/40)	4 (23/621)	0 (0/11)	10 (5/51)	10 (4/40)	10 (256/2635)	0 (0/13)

measures and their impact resulted in inflating the improvement rates in the final sample. Additionally, because of a lack of data on income and deprivation we were unable to assess the potential impact of these factors on outcomes. Only further studies with more complete datasets will be able to address these questions empirically. However, findings from other studies with reduced questionnaire response rates would suggest that, generally, the latter hypothesis holds; i.e. the sample is likely to be skewed to those who feel more positive about their treatment.[37]

The focus on case closure as the only follow-up time point for assessing the impact is also a limitation. There is an increasing interest in the longer-term outcome of therapy and it may be that an impact does not show itself until some years later, as a 'sleeper' effect.[38] However, even if this were the case it would be important to know what the likely outcome is at case closure to support informed expectations around this.

The fact that the research focused on adolescent self-report alone may also be a limitation. It is known that there are significant differences between adolescents, their parents and those seeking to treat them in terms of their perspectives on the nature and the progress of the problem. We chose to focus on adolescent self-report as this viewpoint has been more overlooked in previous research and the experience of the person is clearly crucial. However, future research could also focus on practitioner, parent and teacher perspectives.

The fact that no UK norms existed for the measure used is a limitation and could mean that adolescents completed the questionnaires in different ways that may have skewed the data. However, the fact that change in scores was being considered should minimise the impact of this issue to some extent. A further limitation is the focus on reliable change in symptoms as the key metric of outcome. This can be critiqued from two perspectives. First, the concept of reliable change itself is only from a statistical point of view. It relates to the amount of change in scores that makes it unlikely to be because of measurement error alone. But this does not mean it is a significant change from the perspective of the individual, nor might it be reliable in the lay sense of being relied upon to be maintained over time. Moreover, solely focusing on symptoms could be criticised. A more holistic approach focusing on change in other metrics may be more important to the adolescent's life, such as functioning in school, ability to manage their own difficulties or interpersonal relationships. Both these limitations could be addressed by developing further refined metrics and approaches, ideally coproduced with young people themselves, and are issues we are exploring further. However, in the meantime, reliable improvement on symptom change could be argued to be the best and most transparent measure.

Implications

Notwithstanding the limitations, the findings raise important implications for research and practice in relation to meeting the needs of young people with anxiety and depression. Greater exploration of this area with more complete data-sets, refined metrics and multiple perspectives is needed. Future research could look at combinations of outcomes as proxy for recovery and it will be important to consider qualitative findings to explore further the experience of both recovery and non-recovery in treatment. While this research is being developed and the evidence base extended, there are some potential implications for practice that are considered below.

One key implication for practice is the need for a recalibration of what is said to young people and the wider public about the likely immediate measurable outcome of therapy to rebalance the focus on the possibility of therapy helping with a realistic appropriate appraisal of the fact that not everyone will be measurably improved by the end of treatment.

These findings need to be considered in the context of research into spontaneous remission among adolescents for depression and anxiety. Whiteford et al's ground-breaking review[39] of spontaneous remission in depression found that children and adolescents were more likely to achieve remission without treatment than adults, and achieve it faster, and that spontaneous rates of remission for young people could range up to 48%. In Weisz et al's seminal meta-analysis[12] they note the 'probability that a randomly selected youth in the treatment condition would be better off after treatment than a randomly selected youth in the control condition [which was generally a wait list control] was 63% – only moderately better than chance at 50%'.

These, combined with the current findings and other research,[29–33] thus present an increasingly substantial challenge to the current silence on likely measurable change achieved by the time of case closure. Some practical suggestions from the current findings include: those developing guidance publish metrics on reliable improvement rates – and other metrics of outcome as relevant – in their recommendations to the public; service providers setting realistic outcomes with funders (whether insurance companies or public funders) around likely improvement rates for this population and practitioners considering what they say to children, young people and families accessing their services in terms of likely outcome of an episode of care.[40] Clearly this needs to be balanced carefully with the need to instil hope that has been identified as a powerful moderator in therapeutic impact.[41] However, the current findings would suggest that more open conversations about likely reliable improvement rates are necessary, and potentially more focus from the start as to how treatment might enhance adolescent ability to manage ongoing symptoms and difficulties that are still likely to be present at the close of treatment. There may even be therapeutic benefits to this approach.

Although therapists may say they do not want to share reliable improvement rates out of desire to protect the young person and not destroy hope, it is sometimes not clear who is being protected.[42] Young people with experience of service use as adolescents have reported how they have felt a failure as they believed they were the only ones not to improve with treatment.[40] If more realistic expectations are set with a focus on ongoing patient activation and management of likely ongoing symptoms and difficulties, this may actually prove beneficial in terms of outcomes for those young people. It may also help children, young people and their families make more informed choices about treatment and when to stop. However, clearly this is a hypothesis awaiting further empirical study.

Julian Edbrooke-Childs, Lecturer, Evidence Based Practice Unit, Brain Sciences, UCL and Anna Freud National Centre for Children and Families, London, UK;
Miranda Wolpert, Professor, Evidence Based Practice Unit, Brain Sciences, UCL and Anna Freud National Centre for Children and Families, London, UK; **Victoria Zamperoni**, Research Officer, Child Outcomes Research Consortium, London, UK; **Elisa Napoleone**, Research Officer, Child Outcomes Research Consortium, London, UK; **Holly Bear**, PhD student, Evidence Based Practice Unit, Brain Sciences, UCL and Anna Freud National Centre for Children and Families, London, UK

Correspondence: Miranda Wolpert, Anna Freud National Centre for Children and Families, 12–14 Jordan House, London N1 6EB, UK. Email: miranda.wolpert@ucl.ac.uk

First received 23 Oct 2017, final revision 5 Apr 2018, accepted 15 May 2018

Supplementary material

Funding

Acknowledgements

References

1 McMartin SE, Kingsbury M, Dykxhoorn J, Colman I. Time trends in symptoms of mental illness in children and adolescents in Canada. *Can Med Assoc J* 2014; **186**: E672–8.

2 Lessof C, Ross A, Brind R, Bell E, Newton S. *Longitudinal Study of Young People in England Cohort 2: Health and Wellbeing at Wave 2*. TNS BMRB, 2016.

3 McManus S, Bebbington P, Jenkins R, Brugha T. *Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. NHS Digital, 2016.

4 Perou R, Bitsko RH, Blumberg SJ, Pastor P, Ghandour RM, Grfroerer JC, et al. Mental health surveillance among children – United States, 2005–2011. *MMWR Suppl* 2013; **62**: 1–35.

5 The Children's Society. *The Good Childhood Report 2016*. The Children's Society, 2016 (http://www.childrenssociety.org.uk/sites/default/files/pcr090_mainreport_web.pdf).

6 Bor W, Dean AJ, Najman J, Hayatbakhsh R. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Aust N Z J Psychiatry* 2014; **48**: 606–16.

7 Fink E, Patalay P, Sharpe H, Holley S, Deighton J, Wolpert M. Mental health difficulties in early adolescence: a comparison of two cross-sectional studies in England from 2009 to 2014. *J Adolesc Health* 2015; **56**: 502–7.

8 National Institute for Heath and Care Excellence. *Anxiety Disorders*. NICE, 2014 (https://www.nice.org.uk/guidance/qs53/chapter/List-of-quality-statements).

9 National Institute for Heath and Care Excellence. *Depression in Children and Young People: Identification and Management*. NICE, 2015. (https://www.nice.org.uk/guidance/cg28/chapter/1-Recommendations#care-of-all-children-and-young-people-with-depression).

10 Royal College of Psychiatrists. Mental Health and Growing Up Factsheet. *Depression in Children and Young People: Information for Young People*. Royal College of Psychiatrists, 2017. (http://www.rcpsych.ac.uk/healthadvice/parentsandyouthinfo/youngpeople/depressioninyoungpeople.aspx).

11 Anxiety and Depression Association of America. *Children and Teens: Treatment*. Anxiety and Depression Association of America, 2017 (https://adaa.org/living-with-anxiety/children/treatment).

12 Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *Am Psychol* 2017; **72**: 79–117.

13 Jensen SA, Corralejo SM. Measurement issues: large effect sizes do not mean most people get better - clinical significance and the importance of individual results. *Child Adolesc Ment Health* 2017; **22**: 163–6.

14 Wolpert M. Commentary: why measuring clinical change at the individual level is challenging but crucial–commentary on Jensen and Corralejo (2017). *Child Adolesc Ment Health* 2017; **22**: 167–9.

15 Clark DM, Canvin L, Green J, Layard R, Pilling S, Janecka M. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet* 2017; **391**: 679–86.

16 Evans C, Margison F, Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evid Based Ment Health* 1998; **1**: 70–2.

17 Wolpert M, Görzig A, Deighton J, Fugard AJ, Newman R, Ford T. Comparison of indices of clinically meaningful change in child and adolescent mental health services: difference scores, reliable change, crossing clinical thresholds and 'added value'–an exploration using parent rated scores on the SDQ. *Child Adolesc Mental Health* 2015; **20**: 94–101.

18 Wolpert M, Dalzell K, Jacob J, Bloxham J, Barnard M, Karwatzki E, et al. Routine outcome monitoring in child and adolescent mental health in the United Kingdom at the individual and systems levels. In *The Cycle of Excellence: Using Deliberate Practice to Improve Supervision and Training* (eds T Rousmaniere, RK Goodyear, SD Miller, BE Wampold). Wiley-Blackwell, 2017.

19 Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; **59**: 12–9.

20 Wolpert M, Jacob J, Napoleone E, Whale A, Calderon A, Edbrooke-Childs J. *Child- and Parent-Reported Outcomes and Experience from Child and Young People's Mental Health Services 2011–2015*. CAMHS Press, 2016.

21 Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. *Behav Res Ther* 2000; **38**: 835–55.

22 Department of Health. *Governance Arrangements for Research Ethics Committees (GAfREC)*. NHS Health Research Authority, 2015 (http://www.hra.nhs.uk/resources/research-legislation-and-governance/governance-arrangements-for-research-ethics-committees/).

23 Stein DJ, Fineberg NA, Bienvenu OJ, Denys D, Lochner C, Nestadt G, et al. Should OCD be classified as an anxiety disorder in DSM-V? *Depress Anxiety* 2010; **27**: 495–506.

24 Law D, Wolpert M, editors. *Guide to Using Outcomes and Feedback Tools With Children, Young People and Families (2nd edn)*. CAMHS Press, 2014.

25 Kosters MP, Chinapaw MJM, Zwaanswijk M, van der Wal MF, Koot HM. Structure, reliability, and validity of the revised child anxiety and depression scale (RCADS) in a multi-ethnic urban sample of Dutch children. *BMC Psychiatry* 2015; **15**: 132.

26 Chorpita BF, Ebesutani C, Spence SH. Revised Children's Anxiety and epression Scale User's Guide. UCLA Child First, 2015 (https://www.childfirst.ucla.edu/wp-content/uploads/sites/163/2018/03/RCADSUsersGuide20150701.pdf).

27 Canty A, Ripley B. *boot: Bootstap R (S-Plus) Functions. 1.3-20 ed2017. p. R package*. R, 2017.

28 Cohen J. A power-primer. *Psychol Bull* 1992; **112**: 155.

29 Edlund JN, Carlberg G. Psychodynamic psychotherapy with adolescents and young adults: outcome in routine practice. *Clin Child Psychol Psychiatry* 2016; **21**: 66–80.

30 Jónsson H, Thastum M, Arendt K, Juul-Sørensen M. Group cognitive behavioural treatment of youth anxiety in community based clinical practice: clinical significance and benchmarking against efficacy. *J Anxiety Disord* 2015; **35**: 9–18.

31 Biegel GM, Brown KW, Shapiro SL, Schubert CM. Mindfulness-based stress reduction for the treatment of adolescent psychiatric outpatients: a randomized clinical trial. *J Consult Clin Psychol* 2009; **77**: 855–66.

32 Hayes L, Boyd CP, Sewell J. Acceptance and commitment therapy for the treatment of adolescent depression: a pilot study in a psychiatric outpatient setting. *Mindfulness* 2011; **2**: 86–94.

33 Nilsen TS, Handegard BH, Eisemann M, Kvernmo S. Evaluating change in symptomatic and functional level of children and youth with emotional disorders: a naturalistic observation study. *Eur Child Adolesc Psychiatry* 2015; **24**: 1219–31.

34 Seidel JA, Miller SD, Chow DL. Effect size calculations for the clinician: methods and comparability. *Psychother Res* 2014; **24**: 470–84.

35 Royal College of Paediatrics and Child Health. *National Paediatric Diabetes Audit*. RCPCH, 2015.

36 Child Outcomes Research Consortium. *Flawed, Uncertain, Proximate and Sparse (FUPS) Data*. Child Outcomes Research Consortium, n.d. (http://www.corc.uk.net/media/1533/fupsleaflet.pdf).

37 Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Serv Res* 2002; **37**: 1403–17.

38 Neufeld SA, Dunn VJ, Jones PB, Croudace TJ, Goodyer IM. Reduction in adolescent depression after contact with mental health services: a longitudinal cohort study in the UK. *Lancet Psychiatry* 2017; **4**: 120–7.

39 Whiteford HA, Harris MG, McKeon G, Baxter A, Pennell C, Barendregt JJ, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychol Med* 2013; **43**: 1569–85.

40 Wolpert M. Failure is an option. *Lancet Psychiatry* 2016; **3**: 510–2.

41 Wampold B. *The Great Psychotherapy Debate: Models, Methods, and Findings*. Routledge, 2013.

42 Rousmaniere T, Wolpert M. Talking failure in therapy and beyond. *The Psychologist* 2017; **30**: 40–3 (https://thepsychologist.bps.org.uk/talking-failure-therapy-and-beyond).