Estimating effectiveness of school-based counselling: Using data from controlled trials to predict improvement over non-intervention change

In recent years, there have been growing demands for counsellors and psychotherapists to demonstrate the effectiveness of their interventions (Centre for Mental Health et al., 2012). It is no longer sufficient for therapists to simply say that what they do 'works' (Cooper, 2008): without evidence of positive outcomes, there is an increasing probability that services will lose funding and may become decommissioned (Cooper, 2011). Evidence of effectiveness can also empower potential service users to make informed choices about the kinds of interventions to engage with.

In the field of school-based counselling, as developed and delivered in the UK (i.e., with a particular humanistic or integrative orientation, see Cooper, 2009, 2013; Hill, 2011), some academics and practitioners have responded to this agenda by conducting increasingly large-scale and thorough evaluations of this work. Most recently, two large scale reviews of the outcomes of over 5,000 young people attending school-based counselling in Wales (Cooper, Pybis, Hill, Jones, & Cromarty, 2013; Hill et al., 2011) and the wider UK (Cooper, 2009) have shown, conclusively, that school-based counselling is associated with reductions in psychological distress for young people. The amount of change associated with this intervention (*standardised effect sizes* of 0.81 (Cooper, 2009) and 1.09 (Cooper et al., 2013; Hill et al., 2011)) are consistent with the magnitude of improvement observed in other therapeutic services for young people (CAMHS Outcome Research Consortium, 2012); as well as for adult clients in the UK's National Health Service (NHS) (Clark et al., 2009; Stiles, Barkham, Mellor-Clark, & Connell, 2008).

Such *longitudinal* data, however, in which change is assessed from the beginning to end of an intervention, are not regarded highly in established hierarchies of evidence (Cooper, 2011), and rarely find their way into clinical commissioning guidelines (e.g., National Institute for Health and Clinical Excellence, 2005). A key reason for this is that this data is *uncontrolled.* This means that there is no comparative group of participants, randomly allocated to a non-intervention condition, against which to evaluate the magnitude of change. Hence, while pre-/post-intervention data can show that school-based counselling is *associated* with reductions in psychological distress, it cannot show that the intervention *caused* the reduction in distress, because this reduction may have happened anyway. It may be, for instance, that young people who are experiencing psychological distress simply improve over time, and there is some evidence to suggest that this is, indeed, the case (Daniunaite, Ali, & Cooper, 2012).

It may also be that general improvements over time reflect a 'regression to the mean', whereby more extreme values tend to be closer to the average on their second measurement. This is because questionnaire (and other) scores can be seen as consisting of a 'true' score and a random 'error' component (Lord & Novick, 1968). This random component may be due to many factors, such as moment-to-moment fluctuation in mood resulting from recent events, or the complex process of translating feelings into questionnaire responses. Hence, a particularly extreme score on first measurement (e.g., indicating more distress) is often due to 'chance', and on the next measurement the score will be lower, even if the 'true' score has not changed.

A principal means of identifying the specific causal effects of an intervention is the randomised controlled trial (RCT). Here, the amount of change experienced by individuals who participate in an intervention can be compared against the amount of change experienced by a similar group of participants who do not: a *control group*. This may be people who receive

treatment as usual, who are put on a waiting list for the intervention rather than actually receiving it, or who participate in an alternate intervention to compare effects. Allocation to intervention or control group is random, so that factors causing non-intervention-related improvement in symptoms – many of which may be unknown – are randomly distributed between the two groups. If the intervention is associated with statistically significantly greater amounts of change than the control condition, it can be claimed that the intervention is responsible for bringing about this effect.

In the UK school-based counselling field, four pilot RCTs have now been conducted (Cooper et al., 2010; McArthur, Cooper, & Berdondini, 2013; Pearce, Sewell, & Osman, 2013; Pybis et al., 2014). Overall, they suggest that school-based counselling *is* effective, with clients participating in counselling experiencing significantly greater improvements in wellbeing than young people who are allocated to a waiting list condition (see Cooper 2013). However, the numbers in these studies are relatively small – around 30 participants, in total, per study – and this limits the reliability and validity of the findings. Furthermore, these studies have focused on one specific form of school-based counselling – school-based *humanistic* counselling, drawing from person-centred and experiential principles (Roth, Hill, & Pilling, 2009) – such that these results may not be transferable to other forms of school-based counselling interventions.

Despite these limitations, a key element of these pilot RCTs is that they give some indication of the changes in psychological distress that can be expected for young people who are referred in to school-based counselling in the UK, but who do *not* receive this intervention (in these instances, allocated to a waiting list condition). This raises the question, then, of whether this data could be used more widely, to provide some form of 'benchmark' or 'proxy' control group against which data from much larger longitudinal studies can then be compared.

In other words, if a researcher or school-based counsellor can show, not only that their clients have improved, but improved more than similar young people who do not receive counselling, then this adds considerably to their evidence for effectiveness.

Such a system has been developed for the parent-rated version of the Strengths and Difficulties Questionnaires (SDQ; Goodman, Meltzer, & Bailey, 1998) using data from a community sample of 609 children who exhibited mental health difficulties, but who, primarily (84%), did not receive any treatment for their mental health problems. Data were collected at baseline and then at six months (Goodman, 2007; Ford et al., 2009), such that a formula for predicting changes in untreated children could be established. This predicted score could then be compared with the actual scores observed in children receiving an intervention, to estimate the benefit of treatment versus no treatment. The authors termed this an *added-value score* (AVS), based on the 'value-added' score used in education (e.g., Sanders & Horn, 1994). The Parent-SDQ AVS is routinely used by the CAMHS Outcomes Research Consortium's (CORC) analyses of mental health services' outcomes (Wolpert et al., 2012). It has also been validated in an RCT of parent training for three and four year olds (Ford, et al., 2009), which found that the AVS in the control group, as expected, was close to zero; while the AVS in the intervention group was similar to the overall intervention effect.

The Parent-SDQ AVS is a significant development in the field of outcome evaluation, and the principles underlying this approach can be extended to other measures and clinical contexts. In this paper, we do so with specific reference to counselling in UK secondary schools; and in this regard have adapted Goodman's (2007) original procedures in several ways. First, we have developed formula for the Young Person's CORE (YP-CORE), rather than the SDQ, as this is the principal outcome measure used in this context (Cooper, 2013). Second, we have used

data from young people who were referred to school-based counselling, but who did not receive the intervention, as this was considered the closest approximation to a typical school-based counselling population. Third, we looked at changes for this group at six and 12 weeks from baseline assessment (rather than four and eight months, as in the Goodman study), as this more accurately represents the typical time period of young people in school-based counselling (Cooper, 2013). Finally, we used this study as an opportunity to develop some new terminology for this approach, as we felt the term 'added value' was somewhat ambiguous.

The aims of the present study, therefore, were to develop and test a method of predicting outcomes for young people were they not to receive school-based counselling, such that more accurate indicators of treatment effectiveness could be established for this intervention. We also aimed to present this method in a way that could be adopted by other practitioners and researchers in the field.

Method

Summary of design

This study consists of two phases: calculating a formula for estimating non-intervention outcomes, and then evaluating the utility of this formula with a practice-based sample. In the first phase, data from participants in the control conditions (waiting list) of four previous RCTs of school-based counselling was re-analysed using regression modelling to create a formula for calculating *estimated non-intervention outcomes* (ENOs) for clients in school-based counselling. To test out this formula, in the second phase, ENOs were calculated for participants in a recent evaluation study of school-based counselling, and compared against the participants' *actual outcomes* (AOs), to give *estimated intervention effects* (EIEs). We then analysed the EIEs descriptively and inferentially to assess and illustrate the value of this method.

Phase one: Calculation of estimated non-intervention outcomes

Participants.

To develop our formula for calculating ENOs, we drew on data from 83 young people who had been allocated to the waiting list control conditions of four pilot RCTs of school-based counselling for psychological distress, and for whom there was a baseline score at assessment (see Figure 1). These 83 young people represented 93.3% of those allocated to waiting list conditions in these four trials (n = 89), with 42.9% of those originally assessed (n = 132 of 308) excluded from randomisation, primarily due to insufficiently high levels of emotional distress. Fifteen of these 83 young people were from the SCOPED study (Cooper et al., 2010); 17 young people were from McArthur, Cooper and Berdondini's (2013) pilot trial (termed 'SUPPORT'); 21 young people were from the RELY study (Pybis et al., 2014), and 30 young people were from the ALIGN pilot trial (Pearce, Sewell, & Osman, 2013). Nineteen of these participants (22.9%) were male, 62 participants (74.7%) were female, and gender data was missing for two participants (2.4%). There were no statistically significant differences in gender ratios between the studies, $\chi^2(3) = 5.5$, $p = .14$. The ages of the participants ranged from 11 to 18, with a mean of 14.3 (SD = 1.3) and no significant differences across studies, $F(3,77) = 1.9$, $p = .14$. Data on ethnicity were available for 64 of the participants (77.1%), with 41 of these categorised as white (64%), seven categorised as mixed background (11%), and 16 as black and ethnic minority (25%).

[Insert Figure 1 about here]

All participants had been experiencing moderate or more severe levels of emotional distress at the time of allocation, with an inclusion criterion of four or greater on the Emotional Symptoms scale of the Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998) for the SCOPED study, and five or greater for the other three studies. The mean YP-CORE score at baseline of 19.2 (*SD* = 7.2) indicated that these participants had similar levels of psychological distress at assessment to those typically found in school-based counselling services: for instance, 18.7 for clients from 33 datasets in the evaluation of school-based counselling in Wales (Cooper, Pybis, Hill, Jones, & Cromarty, 2012).

Measures.

The Young person's CORE (YP-CORE, Twigg et al., 2009) was the principal measure used in both phases of this study. It is a ten item, pan-theoretical self-report measure of psychological distress in young people aged 11–16. It was adapted from the CORE-OM ("Clinical Outcomes in Routine Evaluation -- Outcome Measure", Barkham et al., 2001), and is the third revision of this measure (previous versions being a 14-item "Teen-CORE" and an 18-item "Young People's CORE v.1"). The measure asks young people to rate how they have been feeling over the past week on a five-point scale (0 = *Not at all*, 4 = *Most or all of the time*), for example 'I've felt edgy or nervous'. Ratings are averaged and then multiplied by 10, to give an overall score from 0 to 40, with higher scores indicating higher levels of psychological distress. For the purposes of the present study, YP-CORE forms were considered acceptable if eight or more items had been completed. Prorating took place for one baseline score only.

The ten item YP-CORE has been shown to be acceptable to young people; with good inter-item reliability (Cronbach's $\alpha = 0.85$) – consistent across gender and age groups (Twigg et al., 2009). Waiting list data from the first phase of the current study indicated good levels of

inter-item reliability, with a Cronbach's α of .83 at six weeks and .80 at 12 weeks. Data from the second phase of the study again indicated that the measure had acceptable levels of inter-item reliability, with a Cronbach's α of .78 at both baseline and endpoint.

Procedure.

The 15 young people in the SCOPED study had been recruited through a classroom-based screening process, in which they had been asked whether they had problems they would like to talk to a counsellor about. For the 68 young people in the SUPPORT, RELY and ALIGN studies, recruitment had been more representative of real world circumstances, with young people referred into the study by pastoral care teachers who believed that they were experiencing psychological difficulties and might benefit from counselling. All participants had then been assessed by a researcher (*baseline* assessment), and subsequently informed that they would be placed on a waiting list for counselling, either for six weeks (SCOPED) or one school term (SUPPORT, RELY and ALIGN). Informed consent was obtained from the young people directly; but parents and carers were informed (with the young person's consent) that the young person was participating, and were given the opportunity to opt out of the studies on the young person's behalf. Participants were not offered any formal therapeutic intervention during this time period, though they were reminded that they had access to the school's usual pastoral care provision. In addition, in the SUPPORT, RELY and ALIGN studies, participants had mid-point assessment meetings with researchers at approximately six weeks from baseline, in which they completed a number of forms to indicate their levels of psychological distress and wellbeing.

At six weeks post-baseline assessment (Time 2, T2), data were available for 76 of the 83 participants (91.6%). This was endpoint for participants in the SCOPED study, and midpoint for participants in the SUPPORT, RELY and ALIGN studies. At 12 weeks post-baseline (Time 3,

T3), data were available for 62 of the 83 participants in the SUPPORT, RELY and ALIGN studies (74.7%). Paired data was available for *either* T1-T2, *or* T1-T3, for 79 participants; and for *both* time periods for 59 participants.

Analysis.

The first step in our phase one analysis was a simple descriptive inspection of means of outcome scores at approximately six (T2) and 12 weeks (T3) from baseline assessment (T1) for the waiting list participants.

Analyses to calculate the non-intervention outcomes were performed using the software programme R v3.0.0 (R Core Team, 2013). Linear mixed effects models, fitted by maximum likelihood estimation using the lme4 package (Bates, Maechler & Bolker, 2012), were used to predict outcomes for young people referred to – but not receiving – school-based counselling. These models are generalisations of linear regression and allow more than one observation from each participant to be included in analyses. The predictors tested were baseline YP CORE score, weeks from baseline, gender, age, and study, initially fitting a model with all predictors and removing any which were not statistically significant. Between-participant variation in the outcomes which was not explained by these predictors was also modelled. As outlying or overly-influential data points could bias the estimated ENOs, these were identified using both statistical means (DFBETAS over and Cook's distances over) and visual inspection (Van der Meer, Te Grotenhuis & Pelzer, 2010). Significant predictors of outcomes were retained to give a final formula for calculating ENOs.

Phase two: Evaluating the utility of the ENOs

Having calculated the expected change without intervention from the data from these four studies, the utility of the method was evaluated by applying it to an actual school-based counselling outcome dataset.

Participants.

We used data from a 2009–2010 practice-based evaluation of school-based counselling in Scotland (Cooper, McGinnis, & Carrick, in press), which had obtained a relatively high completion rate: 97.3% of those who were eligible and consenting to participate in the evaluation. Data were available on 256 clients, of whom 153 (59.8%) were female and 103 (40.2%) were male. The mean age of the sample was 13.7 (range 11 to 17), and the mean number of counselling sessions attended was 5.1 (range 2 to 27) – giving a mean of approximately six weeks between first and last session.

The baseline YP-CORE score mean for this sample was 18.3 (SD = 7.3), which was not significantly different from the mean for the phase one participants (t = 1.3, p = .21). Descriptive statistics for this sample (where either T1-T2, or T2-T3, data were analysed), and the phase one samples, are given in Table 1. There was no statistically significant difference in the variances ($F(73, 255) = 0.8$, p = .32), or maximum difference in cumulative distributions, i.e., in the proportion of people who score at or below each possible value of YP-CORE (Kolmogorov-Smirnov D = 0.1, p = .24). Together these results suggest no evidence that the two distributions differ.

[Insert Table 1 here]

Procedure.

The form of counselling that clients in this sample participated in has been termed *school-based humanistic counselling* (Cooper et al., 2010). This is based on person-centred principles and practices, with counsellors using a range of core methods including empathic reflections, the provision of unconditional positive regard, and prompts to help clients explore their experiences and emotions (Mearns & Thorne, 2007; Rogers, 1961). All counsellors were fully qualified and had been trained to diploma level in humanistic therapy; with additional formal training in humanistic counselling with adolescents. All practitioners' case-loads were supervised for a minimum of 90 minutes per month by an experienced humanistic practitioner. However, there was no independent auditing of counselling practice to evaluate adherence to school-based humanistic counselling competences.

Analysis.

To test the validity and utility of the ENOs, we used the formula from the first phase of the study to calculate ENOs for the 256 young people from the 2009-10 school-based counselling service data. From this, we then calculated an estimated intervention effect (EIE) for each client, using the formula EIE = ENO – AO (actual outcome at endpoint). Because higher scores on the YP-CORE indicate greater levels of distress, a positive EIE indicated that clients had experienced greater change that would be predicted had they not received an intervention; and a negative EIE indicates that they have experienced less change than would be predicted in a non-counselling condition. We then examined the EIEs descriptively, and conducted a t-test to see whether the mean EIE was significantly greater than 0. This would indicate that the clients had, on average, experienced a greater amount of change than would be expected had they not participated in the counselling. Finally, to calculate a standardised effect size for the intervention (Cohen's *d*), we divided the mean EIE by a norm YP-CORE standard deviation at baseline of

7.45. This norm was derived by calculating the mean SD at baseline across previous evaluation studies using the YP-CORE (Hill et al., 2011).

Results

Phase one: Calculation of estimated non-intervention outcomes

Four pairs of data were removed from the T1 to T2 analysis because the participants had influential datapoints and biased the mean predictions of the ENO models (described later). A further pair of T1-T2 data was removed because the participant's age and gender was missing, and one additional pair of T1-T2 data was removed due to missing information on weeks from baseline. This left 70 pairs of T1-T2 data included in the final analysis. At T1 to T3, three pairs of data were removed, as above, because the participants had influential datapoints, and a further three pairs of data were removed due to missing information on weeks from baseline. This left 56 pairs of T1-T3 data in the final analysis. In total, 74 of the 79 participants provided either T1-T2, or T1-T3, data.

There were statistically significant correlations between YP-CORE scores at T1 and T2 ($r = 0.66$, $p < .001$), and T1 and T3 ($r = 0.61$, $p < .001$).

The mean YP-CORE scores for paired data points, and mean time between measurements, are shown in Table 2. Across the four studies, waiting list participants' mean scores decreased from 19.2 (SD = 6.6) at baseline to 17.6 (SD = 7.2) at approximately six weeks (a reduction of 1.6 points on the YP-CORE), and 19.8 (SD = 6.1) at baseline to 17.2 (SD = 6.3) at approximately 12 weeks (a reduction of 2.6 points on the YP-CORE).

[INSERT TABLE 2 HERE]

Linear mixed-effects models were next used to estimate ENO. There were no statistically significant effects of weeks from baseline (95% CI [–0.2, 0.1]; log-likelihood ratio [LLR] $\chi^2(1) = 0.36$, $p = .55$). This indicated that, on average, the young people's YP-CORE scores were not related to how long from baseline assessment their scores had been taken. There was no significant association between outcome scores for the waiting list participants and age (95% CI [–1.2, 0.4]; LLR $\chi^2(1) = 0.91$, $p = .34$), or gender (95% CI [–3.8, 1.3] with female as reference level; LLR $\chi^2(1) = 0.99$, $p = .32$). Nor was there an interaction between baseline score and study, indicating no difference between studies in how participants' scores change over time (LLR $\chi^2(3) = 4.78$, $p = .19$). However, there was a main effect of study, accounted for by differences in baseline scores (LLR $\chi^2(3) = 15.64$, $p = .001$), which explained 14.2% of the overall variance in outcomes.

Baseline score was a statistically significant predictor of outcome (mean slope = 0.6, 95% CI [0.5, 0.8]; LLR $\chi^2(1) = 47.58$, $p < .001$). Here, higher levels of distress at baseline were associated with higher levels of distress at outcome, accounting for 34.6% of total outcome variance. The 'intercept' (where the regression line crosses the vertical axis) for the final model was 4.2 (95% CI [1.0, 7.4]), giving an ENO prediction formula of:

$$\text{ENO} = 4.2 + (0.6 \times \text{baseline score})$$

Since the slope for the baseline score, 0.6, is less than 1, the difference between the ENO and baseline increases as the baseline score increases as one would expect by regression to the mean. Hence, for example, young people who came to counselling with YP-CORE scores of 15 at assessment, but who did not receive counselling, would be expected to reduce to an average

score of 13.8, a pre-post difference of 1.2; while young people who had a YP-CORE score of 25 at baseline would be expected to reduce to an average score of 20.2, which is a larger pre-post difference of 4.8. Figure 2(a & b) shows scatterplots of baseline against outcome scores for T1 to T2 and T1 to T3. This shows that the ENO (dashed line) is an adequate model of average change for both pairs of time points and that the relation in the raw data is approximately, though not perfectly, linear (solid curve). Figure 2(c) shows pre-post data from school counselling. Here the curve shows average outcomes are lower than was predicted by the ENO, suggesting that counselling had an impact.

[INSERT FIGURE 2 ABOUT HERE]

Phase two: Evaluating the utility of the ENOs

The mean estimated non-intervention outcome (ENO) for the practice-based sample was 15.9 (*SD* = 4.7). The mean actual outcome (AO) was 9.1 (*SD* = 6.2). A paired sample t-test indicated that the mean AO for the school-based counselling clients was significantly lower than the mean ENO ($t = 17.67$, $p < .001$). This indicates that, on average, the clients in this counselling sample improved significantly more than would be predicted had they not received this intervention. This is illustrated in Figure 3, which shows the actual change from baseline to outcome against the estimated non-intervention outcomes. Figure 2(c) shows a scatterplot of baseline and outcome data with a line showing the expected relationship according to ENO if there were no benefit of counselling. As can be seen, most of the outcome scores are below this line, indicating lower scores than predicted.

INSERT FIGURE 3 HERE

The estimated intervention effects (EIEs) on the YP-CORE for the 256 school-based counselling clients in Scotland ranged from -15.3 to 21.2, with a mean of 6.8 (SD = 6.2), and a median of 6.8. The distribution was normal, with no significant skewness (-.19 [SE = .15]) or kurtosis (.03 [SE = .30]). Based on this average estimated effect, the standardised effect size (Cohen's *d*) for the intervention was 0.91, indicating that there was a large difference between the actual outcomes of clients and the outcomes that would be predicted were they not to receive this intervention. The standardised effect size without adjusting for the estimated non-intervention outcome was 1.24, showing that some of this unadjusted effect may be attributed to spontaneous recovery and regression to the mean.

Discussion

Through this analysis of RCT control group data, we developed a formula for calculating the estimated outcomes that young people referred in to school-based counselling would have if they did not receive this intervention: $4.2 + (0.6 \times \textit{baseline score})$. This formula was then tested out in a sample of school-based counselling cohort data, which showed that it could be applied simply and provided useful information about the effectiveness of this intervention over and above estimated non-intervention change. This suggests that the formula developed in this paper can be used more widely by school-based counselling services to evaluate, and present, the effectiveness of their work. This extends work developed using the Parent-SDQ to the YP-CORE measure (Ford, et al., 2009). Such a method can also be used more widely in the counselling and psychotherapy field to help simulate a control group when none is available.

The principal limitation of the method developed and illustrated in this paper is that the control data comes from a population that will inevitably be different from those in the cohort studies for which it is being used as a simulated control. In the present study, for instance, the young people came from specific geographical regions in the UK, with relatively high levels of psychological distress; and were willing to participate in an RCT and for their parents or carers to be told that they were attending counselling. It is possible, then, that young people attending school-based counselling services in other geographical regions, or with other characteristics and features, may change in a non-intervention condition to a different extent. However, the lack of significant differences in means and distributions of baseline scores between our waiting list and school-based counselling sample suggests that our ENO sample may be relatively representative of those attending school-based counselling.

Another important limitation of this study is that the formula for calculating ENOs was based on data only up to 12 weeks, and primarily clustered around either six or 12 weeks, such that the formula may not be applicable for clients outside of this time span. This may be particularly relevant when considering the lack of a significant time effect: were young people to be allocated to a non-intervention condition for a longer period of time, it is quite possible that changes in their levels of psychological distress could be more marked, though it is not clear in which direction this would be.

The relatively small number of control participants also limits the accuracy of the formula for calculating ENOs, and may account for the lack of a time effect. We also used data from a set of studies that was found to be significantly heterogeneous. Another important limitation of this method is that the control participants from whom the ENOs were calculated are not entirely equivalent to a non-intervention condition, as the very act of measuring their

distress and involving them in a trial inevitably affects their levels of distress. However, qualitative studies indicate that such participation is generally experienced as beneficial (Daniunaite et al., 2012), such that the estimation of non-intervention effects derived from such participants is probably greater than would be the case with no intervention at all. This means that the EIEs derived from this formula may be conservative, and more likely to be an under-estimation, rather than an over-estimation, of the actual effects.

To address the non-equivalence of research participants to 'standard' service users, we could have followed Goodman's (2007) procedure of using data from a community sample of young people who exhibited mental health difficulties but who did not receive any intervention. However, as young people in such a sample may not be motivated or intending to attend counselling, they may represent a population more divergent from service users than the present sample.

A final important limitation is that the method adopted in this approach is *not* a randomised controlled design, and is unlikely to be considered of sufficient rigor for national clinical guidelines groups such as NICE.

Despite these limitations, however, the method developed in this paper outlines a means by which practitioners and researchers in the school-based counselling field – and in counselling and psychotherapy more generally – can go beyond describing the amount of change *associated* with their intervention to say something of the *effect* it may be having, at relatively minimal costs. This helps to address a major limitation of cohort- and practice-based data; and may add considerably to the value of evidence at the local stakeholder and commissioning level.

A particular value of the present method is its ease of use, which allows an estimation of effectiveness to be calculated simply through the collection of systematic outcome feedback and

a basic spreadsheet. The key proviso here is that the clients in the evaluation study should be roughly similar to those from whom the formulae have been drawn. For instance, while it may be legitimate to use the present formula for calculating ENOs on the YP-CORE for young people in school-based counselling in the UK, it would be much less appropriate to use with clients outside the UK, or clients in a Tier 4 specialist child and adolescent mental health service.

For practitioners and researchers, the basic steps towards being able to evaluate the effectiveness of an intervention using this method can be summarised as follows:

1. Collect outcome feedback from your clients using the YP-CORE (download from coreims.co.uk). Because a proportion of clients will inevitably drop out, it is important to collect data at every session (unless strongly contraindicated), so that there is a final score for every participant.
2. Calculate the total YP-CORE score for each client at the start of the counselling ('baseline score'), and for the final counselling session ('endpoint score,' or actual outcome [AO]). This can be done by summing the scores on the 10 items at each timepoint (if the client has only completed eight or nine items, divide the total by the number of items completed, and then multiply by ten).
3. Calculate the ENOs for each client by using the formula $ENO = 4.2 + (0.6 \times baseline\ score)$.
4. Calculate the EIEs by using the formula $EIE = ENO - AO$.
5. Calculate a standardised effect size (ES) for the intervention by using the formula: $ES = EIE/7.5$.

6. Calculate whether the estimated intervention effect is significant by using the t-test formula, using two tails and conducting a 'paired' comparison between AOs and ENOs. If the result of this test is a probability of less than .05, it can be claimed that the intervention appears to be bringing about a greater effect than would be expected without the intervention in place.

Given the limitations of this study, this procedure for calculating intervention effects needs to be used with caution. As the term suggests, the EIEs must be understood as *estimations* of intervention effects, and not as a definitive indication of how much change the intervention is bringing about: either at an individual or service-wide level.

Conclusion

Although imperfect, the method developed in this paper shows 'proof of concept', and has the potential to be used more widely for other populations and measures. This means that practitioners and researchers carrying out longitudinal evaluations may be able to make stronger claims for their interventions: not only that it is *associated* with change, but that it is *bringing about* more change than seems likely were an intervention not to be used. At a time when providing evidence of effectiveness is gaining importance to the survival and development of psychological therapeutic practices, this method may be of value to those researchers who are not yet in a position to deliver, or obtain funding for, fully-powered randomised controlled trials.

References

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., . . . McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward

practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*(2), 184-196.

Bates. D., Maechler, M. and Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4

CAMHS Outcome Research Consortium (2012, 13th Nov 2012). [CORC response to member request].

Centre for Mental Health, Department of Health, Mind,
NHS Confederation Mental Health Network, Rethink Mental Illness, & Turning Point. (2012). No health without mental health: Implementation framework. London: Department of Health.

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. [; Research Support, Non-U.S. Gov't]. *Behaviour Research and Therapy, 47*(11), 910-920.

Cooper, M. (2008). *Essential Research Findings in Counselling and Psychotherapy: The Facts are Friendly*. London: Sage.

Cooper, M. (2009). Counselling in UK secondary schools: A comprehensive review of audit and evaluation studies *Counselling and Psychotherapy Research, 9*(3), 137-150.

Cooper, M. (2011). Development of a randomised controlled trial of counselling for depression. Lutterworth: BACP.

Cooper, M., McGinnis, S., & Carrick, L. (in press). School-based humanistic counselling for psychological distress in young people: A cohort study using a low attrition sample. *Counselling and Psychotherapy Research*.

Cooper, M., Pybis, J., Hill, A., Jones, S., & Cromarty, K. (2012). Therapeutic outcomes in the Welsh Government's school-based counselling strategy. *Counselling and Psychotherapy Research*.

Cooper, M., Pybis, J., Hill, A., Jones, S., & Cromarty, K. (2013). Therapeutic outcomes in the Welsh Government's school-based counselling strategy: An evaluation. *Counselling and Psychotherapy Research, 13*(2), 86-97. doi: 10.1080/14733145.2012.713372

Cooper, M., Rowland, N., McArthur, K., Pattison, S., Cromarty, K., & Richards, K. (2010). Randomised controlled trial of school-based humanistic counselling for emotional distress in young people: Feasibility study and preliminary indications of efficacy. *Child and Adolescent Psychiatry and Mental Health, 4*(1), 1-12.

Daniunaite, A., Ahmad Ali, Z., & Cooper, M. (2012). Psychological change in distressed young people who do not receive counselling: does improvement happen anyway? *British Journal of Guidance & Counselling, 40*(5), 515-525. doi: 10.1080/03069885.2012.718734

Ford, T., Hutchings, J., Bywater, T., Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire Added Value Scores: evaluating effectiveness in child mental health interventions. *British Journal of Psychiatry*, *194*(6), 552-558.

Goodman, A. (2007). The Youthinmind formulae for assessing 'value added' by CAMHS treatments: rationale, derivation and preliminary validation. Unpublished note.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry, 7*, 125-130.

Hill, A., Cooper, M., Pybis, J., Cromarty, K., Pattison, S., Spong, S., . . . Maybanks, N. (2011). Evaluation of the Welsh School-based Counselling Strategy. Cardiff: Welsh Government Social Research.

HM Government, & Department of Health. (2011). No health without mental health: A cross-government mental health outcomes strategy for people of all ages. London: Department of Health.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

McArthur, K., Cooper, M., & Berdondini, L. (2013). School-based humanistic counseling for psychological distress in young people: Pilot randomized controlled trial. *Psychotherapy Research, 23*(3), 355-365. doi: http://dx.doi.org/10.1080/10503307.2012.726750

Mearns, D., & Thorne, B. (2007). *Person-Centred Counselling in Action* (3rd ed.). London: Sage.

National Institute for Health and Clinical Excellence. (2005). Depression in children and young people identification and management in primary, community and secondary care *National Clinical Practice Guideline Number 28*. London: NICE.

Pearce, P., Sewell, R., & Osman, S. (2013). *The ALIGN project: A randomised controlled trial of school-based person-centred counselling*. Paper presented at the BACP Research Conference 2013, Birmingham.

Pybis, J., Cooper, M., Hill, A., Cromarty, K., Levensley, R., Murdoch, J., & Turner, N. (2014). Pilot randomised controlled trial of school-based humanistic counselling for psychological distress in young people: Outcomes and methodological reflections. *Counselling and Psychotherapy Research*. doi: 10.1080/14733145.2014.905614

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ .

Rogers, C. R. (1961). *On Becoming a Person: A Therapist's View of Therapy*. London: Constable and Co.

Roth, A., Hill, A., & Pilling, S. (2009). The competences required to deliver effective Humanistic Psychological Therapies London: University College London.

Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: replication in a larger sample. *Psychological Medicine, 38*(5), 677-688.

Twigg, E., Barkham, M., Bewick, B. M., Mulhern, B., Connell, J., & Cooper, M. (2009). The Young Person's CORE: Development of a brief outcome measure for young people. *Counselling and Psychotherapy Research, 9*(3), 160-168.

Van der Meer, T., Te Grotenhuis, M., & Pelzer, B. (2010). Influential Cases in Multilevel Modeling: A Methodological Comment. *American Sociological Review*, *75*(1), 173–178. doi:10.1177/0003122409359166.

Wolpert, M., Ford, T., Law, D., Trustam, E., Deighton, J., Flannery, H., & Fugard, A. J. B. (2012). Patient reported outcomes in child and adolescent mental health services (CAMHS): Use of idiographic and standardized measures. *Journal of Mental Health*, *21*, 165–173.