Module:
Psychological Foundations of Mental Health

Week 5
Psychological therapies: from behaviour modification to behaviour therapy

Topic 4
Evaluating the efficacy of cognitive therapy – Part 2 of 3

Professor Richard Brown
Department of Psychology, King's College London

Lecture transcript

Slide 3
Having looked at a couple of individual trials, let's look, now, at another type of evidence-- the systematic review.

This is a research method used to collate and summarise available evidence on a particular subject. The findings can allow us to make evidence-based recommendations about what treatments might be useful in a particular situation, as well as identify gaps and shortcomings in our current evidence. Finally, by combining evidence, we may identify novel relationships that suggest new hypotheses that can be explored in future studies.

Here are a few essential elements of a systematic review. First, the review needs a clear focus that can be specified as one or more specific questions. For example, the question, does CBT work, would not be sufficient. It's both too broad, but also fails to define what is actually meant by "work."

A better question for systematic review would be, how effective is CBT for depression in adults compared to antidepressant medication? Having a clear question then permits the reviewer to specify a set of criteria that further sets the scope of their assessment. They might limit themselves to trials of conventional face-to-face CBT, or include all forms of CBT delivery.

They might include all adults, or only those under the age of 65. The more restrictions, the more homogeneous the studies will be, and potentially the more robust the findings. However, restrictive criteria will also reduce the number of eligible trials, and limit the generalizability of the results.

The eligibility criteria are then converted into standard search terms to be used when searching the various databases of published studies to identify relevant papers for review and analysis. This process of review, data extraction, and analysis has a well-defined methodology. This increases the reliability of the review process itself, and minimises bias on the part of the reviewer.

The typical review process will include an initial examination of the titles and abstracts of all papers identified by the initial literature searches, resulting in a shorter list of relevant papers that are then read in detail. Descriptions of the study methods and results are extracted using standardised methods. This process is often undertaken by two people to increase reliability.

Transcripts by 3Playmedia

An important part of this process is to assess the quality of the study, again using systematic methods and predefined criteria. We will look at an example in a moment.

This quality assessment is crucial. The aim of the systematic review is to synthesise the evidence and arrive at answers to the starting question. The review should attach more importance, or weight, to the outcome of a high-quality study than a low-quality one. Indeed, one high-quality study showing a particular result may be more important than five low-quality ones showing the opposite. Finally, these results are presented in a systematic way, often in tables that summarise the details of the study and their findings, or graphically.

Slide 4
Systematic reviews may consider all types of evidence, but typically restrict themselves to methods from the top half of the evidence hierarchy that we looked at earlier. RCTs tend to provide higher-quality evidence than observational studies, and these, in turn, better-quality evidence than case series or case reports.

However, as we mentioned, RCTs can themselves be of higher or lower quality, as can observational studies. Methodological shortcomings in RCTs can result in their quality being downgraded, while some observational studies can be upgraded. For this reason, a critical part of the systematic review process is assessing the quality of the evidence in individual studies before drawing any conclusions, or making recommendations.

Slide 5
Here is a commonly-used system to assess quality called grade. This assigns evidence a rating of 4 for the highest-quality study, to 1, that is, very low. A score of 0 means that the study should be disregarded completely.

While a well-conducted RCT would get a quality score of 4, a poor quality RCT may be no more valuable than a single case study. Conversely, an excellent controlled observation cohort study may offer the highest quality of evidence. Some feasibility and pilot RCTs, although useful to prepare for a full-scale trial, are typically down-graded for the evidence that they can offer when evaluating efficacy.

Slide 6
Various characteristics of the study, and how it was reported, are used to decide whether a study's quality rating should be upgraded or downgraded. An RCT may be downgraded if there are problems with maintaining blindness and allocation in determining how well the study retained participants into the trial, and avoiding withdrawals in terms of the amount of incomplete or missing data, and in terms of other methodological concerns such as choice of outcome measure.

Conversely, an observational study may be upgraded if it was designed to minimise potential bias, and showed a large treatment effect. Another positive factor would be if a study showed a robust association between the dose of a treatment and the clinical outcome.

In the case of CBT for depression, an observational study might show that the more therapy sessions a patient received, the better their outcome. This does not prove that the therapy is effective, but would be considered strong supportive evidence.

Slide 7
Let's turn next to the quantitative element, included in many systematic reviews of treatments-- the meta-analysis. Meta-analysis is a statistical approach that allows us to combine the findings from a number of different studies to draw overall conclusions about a treatment's efficacy. The advantage of this method is it allows us to pool a number of smaller studies that, individually, may provide inconclusive evidence, but which together can provide a clearer picture, and allow us to estimate the

treatment effect.

With a large number of small trials published in psychotherapy, the meta-analysis is a key tool to help us assess the effectiveness of the treatment and guide future research. The number of meta-analyses has growing steadily over the years. A review of CBT for all treatment indications, published in 2013 by Hoffman and colleagues, showed 269 separate meta-analyses published between the years 2000 and 2011.

Slide 8
When evaluating results from different studies, we often end up having to compare apples and oranges. Obviously, we need a way of describing the results of the different studies in a standard way, before we can combine them in a meta-analysis.

This is a so-called standardised effect size. There are many different ways of calculating such effect sizes, but we will look at two or more of the common ones, here, used depending on the type of data. First, let's consider continuous data-- that is, outcomes that are assessed on a continuous scale, like scores on a depression rating scale.

We are interested in the size of the change in average symptoms following treatment in the active condition, compared to the controlled condition. This graph shows the results of two treatments, A and B, before and after therapy. Each group shows a reduction in the mean depression score after treatment.

The mean change for group B can then be subtracted from the mean change in group A, and an overall difference shown. The raw difference scores are standardised by dividing the mean difference by the standard deviation of all participants at baseline.

A score of 0 indicates no difference between the change following treatment for the two groups. A positive score indicates that treatment favours group B, while a negative score favours group A.

Standardised different scores are calculated in slightly different ways across studies, but typically use one or two statistics, called Cohen's d index, or the Hedges g. In both, a standardised mean difference of 0.8 or above is often regarded as large, but a 0.5 moderate, and 0.2 small.

Slide 9
For some trials, rather than reporting the change in severity of depression symptoms, the outcome of interest is a discrete event such as recovery. Depending on the criteria used, a patient will either have recovered, or not, by the end.

Other trials may look at prevention of relapse. The standardised measure used in such cases is typically the relative risk ratio. This is the relative likelihood that a patient will show with the events outcome in the active group, relative to the control group. Depending on whether the outcome is defined in a positive way-- for example, recovery-- or a negative way-- for example, relapse-- relative risk with either be higher or lower in the effective treatment.

So for example, if 30% of patients receiving the active treatment recovered by the end of the trial, and only 10% of those in the control group, the risk ratio would be 3.0. Risk ratios are also standardised with a ratio of 1 indicating no differential risk.

Slide 10
We will not describe the process of how to carry out a meta-analysis. That's beyond the scope of this module. Instead, we will look at how the results are typically presented, such as the table shown here.

You need to know how to read such tables to be able to understand and interpret the results of meta-analyses. Such tables vary in complexity, and the precise nature of the information presented.

Transcripts by 3Playmedia

However, this one can be used as an example. You'll find more information in this week's reading list.

The first column, on the left, identifies the studies that have been included in the analysis, following the systematic review. The next four columns provide some data about the active intervention group and the control group for each study. In this example, what is being assessed is whether a patient showed a particular outcome following treatment.

The columns marked "events" show the number, and the intervention, and the control group that show this outcome. An example outcome event may be remission or relapse. Other studies use continuous measures, such as scores on the Hamilton depression rating scale.

The next columns, marked "total," show the number of patients in each group receiving treatment. The column marked "weight" describes the proportion of the total number of participants in the analysis contributed by that study. The larger the number, the higher the proportion. We see, here, that Study 6 contributed 46.4% of the total number of participants, so carries the highest weight in the overall analysis.

The next column in the example shows the standardised relative risk statistic for each study, plus its 95% confidence interval. This confidence interval is an indication of how much we should trust that the estimates of the treatment effect measured by the single study is a true reflection of the actual effect. The wider the range, the less our confidence. A 95% confidence interval describes the range within which we would expect the true result to fall if we were to repeat the trial 95 times.

The final column shows a graphical representation of the relative risk ratios, called a forest plot. This shows the distribution of statistics around the midpoint of 1, which indicates no difference between the two groups.

In this example, a low relative risk is the desired outcome. The relative risk is shown by the blue square for each study. We see that all of the studies suggest a positive treatment effect.

We also see two lines on either side of the square. These are the 95% confidence intervals. The wider they are, the less confident we are in the results. Here, the widest confidence interval was shown by Study 5, and the smallest by Study 6.

Typically, if one of the bars overlaps the midpoint line, the study will not have been statistically significant. In this example, none of the studies individually produced significant results.

This is where meta-analyses prove useful. They allow us to not only summarise individual studies in a systematic way, but we can pool the evidence provided to give an overall estimate of the treatment effect. This is shown by the black diamond shape on the bottom of the table.

The centre of the shape shows the overall relative risk, and the width, the pooled confidence interval. We see, here, that there is some support for the treatment effect, but that the diamond still just crosses the midline, suggesting that the effect is not significant-- something is supported by the statistics presented at the bottom.

One other feature to notice about the forest plots is the different sizes of the blue squares. This is a reflection of the weight assigned in the summary analysis. One of the main determinants of quality is often the sample size, as we see here.

Slide 11

Let's return to considering the evidence for CBT for treating depression in adults. This table and forest plot comes from a systematic view and meta-analysis published in 2009 based on available evidence at that time. It comes from the guidance published by the National Institute for Health and Clinical Excellence, or NIHCE, in England.

The full report contains a highly-detailed assessment of all the different types of trials assessing treatment for depression in adults, both drug treatments, psychological treatments, and others. We see, here, the analysis of studies that compared CBT with antidepressant medication. Here, CBT is classed as the active treatment, and medication as the control.

As we shall see later, depression severity can have an effect on treatment outcomes. So this table has split the results into three levels of depression severity. This analysis is based on the difference in improvements between the two groups, the standardised mean difference, or SMD, on the Hamilton Depression Rating Scale.

We see the results from 1,417 patients across 12 individual studies published between 1984 and 2008. We can see, quite clearly, both from the individual studies and the overall means, that there is robust evidence that CBT for depression, in all of these groups, is as effective as antidepressant medication, at least by the end of the treatment period.

Slide 12
Here's another comparison that addressed a slightly different pragmatic question. Given that many patients with depression are treated with antidepressants, is there any advantage in combining medication and CBT, compared to medication alone? This meta-analysis also considers trials published up to 2009.

We see that most of the individual studies were small. And although the results tended to favour the benefits of combined treatment, none showed a clear statistical advantage. The one exception was the largest study, with almost 450 participants, by Keller, published in 2000. This-- showed a significant unreliable advantage of adding CBT to antidepressant treatment. The weighting of this study means that the overall meta-analysis favoured combined treatments in relation to change in depression severity, post-treatment.

Slide 13
Since that meta-analysis in 2009, another important study has been reported. This one was published in 2014 by Steve Hollon and colleagues. It was a large study with 452 adults recruited and treated in three centres. Importantly, these were patients specifically selected for having chronic depression-- an important group for whom existing treatments had not produced any significant or lasting benefit.

Unlike the majority of trials previously, treatment was not stopped after a fixed amount of time, or number of CBT sessions. Instead, it continued for up to 3 and 1/2 years, or until the patient showed a period of sustained remission. This was defined as a Hamilton depression score of less than 9 for a period of four consecutive weeks. The final clinical endpoint was recovery, which was defined as a period of six months following remission, without relapse.

Because previous evidence had pointed to a difference in treatment response between those with more and less severe depression, equal numbers of the two were recruited for the study to allow a second comparison, one based on depression severity. Although not a conventional trial, it was randomised. Standard care-- in other words medication alone-- was the control condition, and medication with adjunctive or additional CBT, the active treatment. Although the patients were not blind to the treatment they were receiving, those doing the assessments were.

Slide 14
What we see, here, are the same results for the low- and high-severity groups, in the medication and CBT plus medication conditions. These sorts of graphs show the cumulative proportion of patients have recovered over time.

The dotted lines and arrows point to the time from the start of treatment when half of the patients have recovered. This is the median time to recovery. On the left, we see the results for the low-

Transcripts by 3Playmedia

severity group. The median time to recovery for the combined condition was 15 months, slightly earlier than for the medication alone group.

This difference was not statistically significant. We can also see, from the two curves, the gradual recovery of the two groups was very similar, with around 70% showing sustained remission by the end of the trial at 42 months.

A different picture is seen on the right for the group with high depression. The median recovery rate for those receiving the combined treatment, at 17 months, was only slightly longer than for the low severity group. However, those receiving medication alone took six months or longer before half had recovered.

Finally, and overall, almost 80% of the more severe patients recovered with a combination therapy, compared to around 60% of those receiving medication alone. This is a substantial difference. More patients are showing sustained recovery and they are recovering sooner.

Slide 15
This set of graphs addresses a second question-- is there any difference between those who have had a period of sustained depression, lasting over two years before the trial, compared to those with equally severe symptoms, but for a shorter time? What we see is that the advantage of combining CBT with medication is absent in those with a long duration of depression. In contrast, there was dramatic advantage in those with a shorter period of depression prior to combined treatment.

What do these results tell us? The most important point is that there is no simple answer to the question of whether combining CBT with medication leads to faster and more sustained recovery. The results indicate that for those patients with either chronic depression, or those with less severe symptoms, there was no added benefit in adding CBT treatment.

However, for patients with severe symptoms, or those with depression that had been present for less time, there were substantial benefits. We might speculate, but cannot provide, answers from these analysis that patients with severe but short-duration symptoms might do best of all.

Large and well-conducted trials such as this offer the opportunity to identify those groups of patients that will benefit from one treatment approach rather than another. Such evidence is critical in deciding how we plan and utilise scarce health service resources, targeting treatments at those patients most likely to benefit.

Slide 16
So we have seen there is evidence that CBT can be an effective treatment, either on its own or in combination with medication. However, before we leave this section, there was one more important question that we can ask. If CBT is an effective treatment for depression, is there something special about the approach? In other words, is CBT more effective than other psychotherapies that are not based on the cognitive model, but on other therapeutic principles and rationales?

We saw, earlier, from the Elkin study, that there was no evidence in that one trial for any superiority of CBT over the other therapy approaches assessed. However, that was only a single study. What would a meta-analysis show?

This one, by Pim Cuijpers, published in 2013, analyses trials between 1966 and 2011. We see, here, the range of alternative therapies evaluated alongside CBT.

The most commonly used was so-called non-directive supportive therapy. This is not a specific therapy type, but an attempt, in the context of a clinical trial, to offer a control condition that has many of the features of cognitive therapy, but without the specific cognitive approaches.

As the name implies, the therapist is supportive, but does not offer any specific guidance on how to understand and manage the symptoms of depression, or the problems that may be contributing to it. The use of such an approach meets some of the needs of a placebo comparator.

Slide 17
Here are the results from the meta-analysis. It's presented in a different way to what we have looked at so far. It shows only the overall findings, without the individual studies, and does not include the forest plots.

The key statistic, here, is called Hedges g, or g in the table. This is the standardised mean difference that we saw previously, along with the 95% confidence interval. It is the statistic that we saw represented by the black diamonds in the forest plots, earlier. A positive g, here, indicates that the results favour cognitive therapy, and negative ones that the results favour the alternative therapy.

What we see is that in none of the comparisons made was there a significant benefit of cognitive therapy, or CBT, compared to an alternative type of therapy, including non-directive supportive therapy. In all cases, the effect size of the difference was very small, and only even approached significance in the case of the comparison between CBT and psychodynamic psychotherapy.

One point to note, however, is that the majority of these studies had methodological weaknesses that limited their quality. First, it does not mean either that CBT doesn't work, or that there is no difference to placebo, although that is one plausible conclusion. Rather, it suggests that CBT, based on the cognitive model, may not be the only way to achieve symptomatic improvement using a psychological approach.

Alternatively, or in addition, it may imply that all therapies share common nonspecific factors that drive symptomatic improvement.