The Cost-Effectiveness of Different Formats for Delivery of Cognitive Behavioural Therapy for Depression: A Systematic Review Based Economic Model

Running title: Cost-Effectiveness Model for CBT delivery modes

Qi Wu, MSc,[1]

Jinshuo Li, M.Phil,[1]

Steve Parrott, MSc,[1]

José Antonio López López, PhD,[2]

Sarah R Davies, PhD,[3]

Deborah M Caldwell, PhD,[4]

Rachel C Churchill, PhD, [5]

Tim J Peters, PhD,[4]

Glyn Lewis, FRCPsych, PhD,[6]

Debbie Tallon, MSc,[7]

Sarah Dawson, MSc,[4]

Abigail Taylor, BM, BCh,[7]

David S Kessler, MD,[7]

Nicola Wiles, PhD,[7]

Nicky J Welton, PhD,[4]

Authors' Affiliations:

[1] Department of Health Sciences, University of York, Helsington, York, United Kingdom

[2] Department of Basic Psychology & Methodology, University of Murcia, Murcia, Spain

[3] School for Policy Studies, University of Bristol, Bristol, UK

[4] Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK

[5] Centre for Reviews and Dissemination, University of York, York, UK

[6] Division of Psychiatry, Faculty of Brain Sciences, University College London, UK

[7] Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

Correspondence to: Ms. Qi Wu, MSc, Mental health and Addiction Research Group, Department of Health, University of York. Heslington, York, UK, YO10. Email: qi.wu@york.ac.uk. Phone: +44 (0) 1904 321 819. Fax: +44 (0) 1904 321 383.

Author Contributions:

Concept and design: Wu, Li, Parrott, Davies, Churchill, Lewis, Taylor, Wiles, Welton

Acquisition of data: Wu, Li, López-López, Davies, Caldwell, Churchill, Tallon, Dawson, Taylor, Kessler

Analysis and interpretations of data: Wu, Li, Parrott, López-López, Caldwell, Churchill, Peters, Wiles, Welton

Drafting of the manuscript: Wu, Li, Parrott, Tallon, Kessler, Wiles, Welton

Critical revision of the paper for important intellectual content: Wu, Li, Parrott, López-López, Davies, Peters, Lewis, Tallon, Taylor, Kessler, Wiles, Welton

Statistical analysis: Welton

Provision of study materials or patients: Caldwell, Lewis,

Obtaining funding: Churchill, Peters, Lewis, Wiles, Welton

Administrative, technical, or logistic support: Caldwell, Tallon, Dawson

Supervision: Parrott, Churchill, Welton

Other: Dawson (*Literature searches)*

Conflict of Interest Disclosures:

Ms. Wu reports grants from National Institute for Health Research (RP-PG-0514-20012), during the conduct of the study. Ms. Li reports grants from National Institute for Health Research (RP-PG-0514-20012), during the conduct of the study. Steve Parrott reports a grant from the NIHR Programme Grants for Applied Research, which funded the original study. Dr. López-López reports grants from NIHR, during the conduct of the study. Prof Churchill

reports grants from National Institute for Health Research (Programme Grants for Applied Research), during the conduct of the study. Dr. Peters reports grants from the NIHR Programme Grant for Applied Research scheme during the conduct of the INTERACT study. Dr. Lewis reports grants from UCL, during the conduct of the study. Dr. Wiles reports grants from NIHR, during the conduct of the study. Dr. Welton reports grants from Pfizer Ltd, outside the submitted work. No other disclosures were reported.

Funding/Support:

This study is funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research (Integrated therapist and online CBT for depression in primary care, RP-PG-0514-20012). This study was also supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol.

Role of the Funder/Sponsor:

The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements

We are grateful to a number of colleagues who are involved with the INTERACT study as co-applicants but who have not participated in drafting this manuscript: David Coyle, Simon Gilbody, Paul Lanham, Una Macleod, Irwin Nazareth, Roz Shafran, Katrina Turner, Catherine Wevill, and Chris Williams.

This study was conducted in collaboration with the Bristol Randomised Trials Collaboration (BRTC), a UKCRC Registered Clinical Trials Unit (CTU) which, as part of the Bristol Trials Centre, is in receipt of National Institute for Health Research CTU support funding.

Word Count: 4,321

Number of pages: 42

Number of Figures: 3

Number of Tables: 3

Appendix:

Pages: 18

Figures: 2

Tables: 7

Abstract

Objectives

Cognitive behavioural therapy (CBT) is an effective treatment for depression. Different CBT delivery formats (face-to-face (F2F), multimedia (MM) and hybrid) and intensities have been used to expand access to the treatment. The aim of this study is to estimate the long-term cost-effectiveness of different CBT delivery modes.

Methods

A decision-analytic model was developed to evaluate the cost-effectiveness of different CBT delivery modes and variations in intensity in comparison with treatment-as-usual (TAU). The model covered an average treatment period of four-month with a 5-year follow-up period. The model was populated using a systematic review of randomised controlled trials and various sources from the literature.

Results

Incremental cost-effectiveness ratios (ICERs) of treatments compared with the next best option after excluding all the dominated and extended dominated options are: £209/QALY for 6(sessions)×30(minutes) F2F-CBT vs TAU; £4,453/QALY for 8×30 F2F vs 6×30 F2F, £12,216/QALY for 8×60 F2F vs 8×30 F2F; £43,072/QALY for 16x60 F2F vs 8×60 F2F. 8×30 F2F-CBT has the highest net monetary benefit for thresholds of £20,000-£30,000/QALY. Probabilistic sensitivity analysis illustrated 6×30 F2F-CBT had the highest probability (32.8%) of being cost-effective at £20,000/QALY, 16×60 F2F-CBT had the highest probability (31.0%) at £30,000/QALY.

Conclusions

All CBT delivery modes on top of TAU were found to be more cost-effective than TAU alone. Four F2F-CBT options (6×30, 8×30, 8×60, 16×60) are on the cost-effectiveness

frontier. F2F-CBT with intensities of 6×30 and 16×60 had the highest probabilities of being cost-effective. However, the results should be interpreted with caution due to the high level of uncertainty.

Highlights

Depression is a common mental health problem associated with a substantial reduction in quality of life. Cognitive behavioural therapy (CBT) is an effective psychological treatment for people with depression and may have a longer-lasting effect compared with pharmacological treatment alone. Different CBT delivery formats (face-to-face (F2F), multimedia (MM) and hybrid CBT) and intensities have been used to expand access to the treatment.

A decision-analytic model was developed to evaluate the long-term cost-effectiveness of different CBT delivery modes and variations in intensity in comparison with treatment-as-usual (TAU). The model covered an average treatment period of 4-month with a 5-year follow-up period and was populated using a systematic review of randomised controlled trials. This model can be easily adapted for use in future studies, to explore the long-term cost-effectiveness of other depression treatments.

All CBT delivery modes on top of TAU were found to be more cost-effective than TAU alone. Four F2F-CBT options (6×30, 8×30, 8×60, 16×60) are on the cost-effectiveness frontier. F2F-CBT with intensities of 6×30 and 16×60 had the highest probabilities of being cost-effective. However, the results should be interpreted with caution due to the high level of uncertainty.

Introduction

Depression is a common mental health problem associated with a substantial reduction in quality of life [1-3]. It is also a chronic relapsing condition and heightens the risk of suicidal behaviour [4-8]. According to the Adult Psychiatric Morbidity Survey, the prevalence of depression in people aged 16 and above was about 3.3% in England in 2014 [9]. Using population estimates from mid-2017, this equates to around 1.6 million people with depression in the UK in 2017 [10].

Depression is associated with an increased economic burden for both individuals and society. The estimated annual cost of health services for depression in England was £1.7 billion in 2007, with an additional £235 million spent on antidepressant drugs [11-13]. Cognitive behavioural therapy (CBT) has been shown to be an effective psychological treatment for people with depression [14]. CBT has been widely used as an alternative or adjunct to pharmacological treatments alone, with some evidence of a longer-lasting effect compared with drug treatments [15,16]. Despite increased investment in psychological services, it is difficult to meet the demand for CBT [17,18]. In order to expand access to CBT, improve the efficiency of service delivery and reduce waiting time for both therapists and patients, alternative modes of CBT delivery have been developed based on modern technology such as telephone, computer and mobile devices [18-20].

Apart from the delivery modes, the intensity (i.e. amount of treatment/dose) of CBT constitutes another key factor in the use of health resources and may affect treatment outcomes [21,22]. According to the National Institute of Health and Care Excellence (NICE) Clinical Guidelines 90 (CG90), adults with depression should be offered psychological interventions such as CBT [23]. Exactly what constitutes the optimal intensity (duration and frequency) of CBT, however, remains unknown [21,23].

We performed a comprehensive systematic review and network meta-analysis (NMA) to compare the effectiveness of different delivery modes of CBT interventions for adults with a primary diagnosis of depression and for whom CBT was considered by their GP. The NMA included 68 randomised controlled trials (RCTs) using standardised diagnostic criteria such as DSM-III, DSM-IV-TR, DSM-5, ICD-10, or validated depression symptom questionnaires to identify depression were included. The details of the review and the results can be found in a separate paper published in 2019 [24].

The delivery modes included were: traditional face-to-face (F2F-CBT) (either individually or in groups); CBT conducted solely via multimedia (MM) platforms (e.g. self-help books, telephone, audio/video recordings, computer programmes, apps, e-mail); and hybrid-CBT interventions, which involved a mixture of face-to-face sessions and multimedia features. The NMA found that MM and hybrid-CBT might be as effective as F2F-CBT. The review also examined the impact of intensity on the effectiveness of different CBT strategies based on the average number and length of CBT sessions. The aim of the current study is to evaluate the cost-effectiveness of different CBT delivery modes and variations in intensity in addition to treatment-as-usual (TAU) in comparison with TAU alone.

Methods

Model structure

A decision-analytic model was constructed to evaluate the cost-effectiveness of the different CBT options in addition to TAU in comparison with TAU alone. For brevity, CBT options plus TAU will be referred to by their CBT elements. The model has two phases: a decision tree to assess the short-term cost-effectiveness of the various CBT interventions during a four-month treatment period, followed by a Markov Model with a cycle of one month to extrapolate the longer-term cost-effectiveness over the subsequent 5-year period. The time

horizon was selected based on the best available data for long-term follow-ups for CBT and supported based on discussions with clinical experts [24,25].

Figure 1 illustrates the structure of the decision tree. Intervention options included in the model are TAU alone, F2F-CBT, hybrid-CBT and MM-CBT. The intensities of F2F and hybrid-CBT were defined in terms of pre-specified combinations of number and length of sessions delivered by a therapist. Based on clinical opinions and the systematic review, the number of sessions was specified as 6, 8, 10, 12, 14 or 16. The length of each session was specified as either 30 or 60 minutes. This allowed for 24 possible intervention combinations (6(sessions)×30(minutes) F2F, 8(sessions)×60(minutes) hybrid, etc.). MM-CBT consisted of mainly self-guided or minimally supported CBTs, and therefore intensity was not measured.

After patients were provided with CBT, they could either complete the intervention or withdraw from therapy. We do not model withdrawals from TAU, as TAU is included in all intervention options and these effects would therefore cancel out in an incremental analysis. We assumed that patients who completed the CBT intervention would incur the full course costs while those who withdrew would incur zero costs of CBT. At the end of the 4-month treatment period, under all options patients were classified as either: responders in remission, responders without achieving remission, or non-responders. Response was defined as 50% reduction in the Beck Depression Inventory (BDI) from baseline, while remission was defined as a BDI score of less than 10 points, commonly used outcomes in previous depression trials [26,27]. It is assumed that all the patients will receive no further depression treatment after the 4-month treatment period.

Figure 2 illustrates the structure of the Markov model which follows the decision tree. The model and the health-state transitions, including remission and relapse beyond the treatment period, take into account the high risk of relapse or a recurrence of depression even after a

successful treatment [7,28,29]. In the model, each patient can be in one of three mutually exclusive health states: depression, remission or dead.

The proportion of responders in remission at the end of the decision tree was applied to a hypothetical cohort of depressive patients in the initial remission state. Responders who didn't achieve remission (including both responders without achieving remission and non-responders) at the end of the decision tree were classified as being in the depression state. We used a simulated cohort of 10,000 depressive patients, with an average age of 45 based on the trial data where most of the model parameters derived from [26,27]. The cohort enter the Markov model after the end of treatment and could change their health state following the direction of the arrows at the end of each monthly cycle for 60 cycles. The model was programmed and run using Microsoft Excel®2016.

Model parameters

Transition probabilities

The decision tree required estimates of the probabilities of withdrawals, response and remission, response without remission, and non-response at end of treatment. Response and remission probabilities were derived by assuming a bivariate normal distribution for baseline and follow-up depression scores on the BDI scale for each intervention, and then applying the definitions of response and remission to obtain proportion of patient with response with or without remission (Appendix-1). The estimated BDI at baseline and follow-up for TAU was obtained from the control arm of the IPCRESS trial which was felt to best represent contemporary current TAU in the UK setting (Appendix-2, Table A1 & Table A2). Relative effects were obtained by re-analysing the data from the systematic review jointly synthesising data an all depression outcomes reported in the papers to obtain mappings between intervention effects on the different depression scales (Appendix-3, Table A3 & Table A4).

This enabled us to combine all evidence and obtain intervention effects on the BDI scale for use in the economic model.

Intensity of intervention was defined as the product of number of sessions by average session length. We estimated a linear effect of intensity on the BDI scale for F2F and Hybrid-CBT interventions with slope -0.496 (95% CrI: -0.922 to -0.063) indicating a larger reduction with increasing intensity (Appendix-3). Due to lack of evidence on Hybrid CBT, we assumed a common slope for F2F and Hybrid CBT [24].

The NMA found no difference between interventions in withdrawal probability, but there was a high degree of heterogeneity in these figures [24]. We therefore assumed withdrawal to be equal for CBT interventions and fitted a random effects single arm meta-analysis of all CBT arms on the log-odds scale to estimate withdrawal probabilities for the model. The results were summarised in two ways: the mean of the random effects distribution (used in our base-case) and the predictive effect in a new population from the random effects distribution (used in a sensitivity analysis) [30] .

The probability of remission and relapse in the post-treatment Markov model was derived from the long-term follow-up data of the CoBalT study, an RCT of CBT conducted in the UK with the longest follow-up period (average of 46 months) [25]. In the Markov model, based on the long-term follow-up data extracted from the CoBalT trial, instead of estimating transition probabilities for each intensity, we assumed that after the end of treatment, the transition probabilities, as well as costs and utilities, only differ between CBT arms and TAU arms.

Based on the CoBalT trial data, the relapse rates were assumed to differ from the 0-6 months period, to the 7-month-5-years period, following the end of treatment. Each patient has a risk of dying in either the depression or remission state. We assume that patients in remission have the same mortality rate as the general population. The general population mortality rate

was obtained from the Deaths registered in England and Wales in 2016 [31]. The relative risk of mortality in depressed relative to non-depressed people was estimated as 1.52 (95% CI=1.45–1.59) from a meta-analysis that included 293 studies and 1.8 million participants [32].

Resource use and costs

The health economic analysis was carried out from the UK National Health Service (NHS) and personal social services perspectives [33]. All costs are presented in Pound sterling at 2016/17 prices. The resources used during the treatment period were: staff time (including therapist and supervisor time), supplementary materials such as booklets, remote delivery methods such as telephone and internet. We used a middle-point NHS pay grade (Band-7) (£36,612 per year) to account for CBT therapists and Band-8a (£44,310 per year) for supervisors[34]. Taking into account the salary on-costs and overheads, the hourly cost of a Band-7 and Band-8a staff member was estimated at £51 and £62 respectively. For the multimedia component in hybrid and MM-CBT, unit costs of materials such as books and internet are given in Appendix-4, Table A5. They were estimated using the weighted average cost of materials used in the studies identified by the systematic review.

The 5-year post-treatment costs were obtained from the CoBalT trial [25] which collected data on health service resource use for patients at 6, 12 and 46-month follow-ups and derived costs for primary and community care, hospital care and the direct costs of personal social services. Using CoBalT data, we derived the monthly costs for patients who were in remission state and those who were in depression state, respectively.

Utility

The primary outcome measure used in the model was Quality-Adjusted Life Years (QALYs). The instrument recommended by NICE to calculate QALYs is the European Quality of Life-5 Dimensions (EQ-5D) measure [33]. We used data from two large UK based studies (IPCRESS and CoBalT) with a total sample of 766 patients with depression to compute the

EQ-5D utility values for different health states before and after treatment [25,27]. The long-term utility values used in the Markov model were derived solely from the CoBalT follow-up study [25]. QALY estimates were calculated by multiplying the utility value for each health state with the time spent in that state using the area under the curve [35].

Cost-effectiveness analysis and sensitivity analysis

An incremental cost-effectiveness analysis was conducted to establish the value for money of different CBT options over and above TAU. We assumed that all patients in the intervention and control groups received TAU. The difference in mean costs between the interventions and the control was divided by the difference in mean health outcomes (QALYs) to generate incremental cost-effectiveness ratios (ICERs) with a time horizon of 5 years after the end of treatment. Based on the maximum acceptable ICER range of £20,000-£30,000 per QALY gained used by NICE, ICERs below this range suggest that the intervention is considered to be cost-effective compared with the control [33]. Meanwhile, net monetary benefit of the interventions was calculated for thresholds of £20,000 or £30,000 per QALY. All costs and QALYs were discounted to a present value at an annual discount rate of 3.5% [33].

Probabilistic sensitivity analysis (PSA) were performed to evaluate the uncertainty surrounding the results. The model was probabilistic in that all parameters were assigned probability distributions to reflect their sample variability [36]. For the short-term decision tree probabilities estimated using Markov Chain Monte Carlo (MCMC), the simulated MCMC values were sampled directly to propagate uncertainty in the input parameters throughout the model. Random values of input parameters were drawn from the assumed distributions: beta distribution for probabilities and gamma distribution for both QALYs and costs [36]. The expected costs and QALYs for each option were calculated using combinations of parameter values [37]. This process was repeated 10,000 times, and the results of these simulations were reported in the form of cost-effectiveness acceptability curve (CEAC). We performed an

extra sensitivity analysis to test the impact of RR of mortality in the model. An extreme value of RR = 1 was used in the sensitivity analysis, i.e. assuming there is no increased risk of mortality in depressed people compared to non-depressed people.

Results

We identified 65 intervention arms with 2,210 patients who received F2F-CBT, of which 35 comprised group therapy. There were only seven hybrid-CBT arms identified in the systematic review, with 401 participants in total. Another 18 study arms were identified as MM-CBT, involving 1,480 patients. The treatment period was four months, in line with most of the included studies.

Transition probabilities

Table A4 lists the estimated response and remission probability for each comparator at the end of treatment based on Appendices 1-3. The withdrawal probability for CBT interventions estimated as the random effects mean was (0.18, 95% CI: 0.15, 0.21) whereas the predictive distribution was (0.20, 95% CI: 0.04, 0.50) which reflects the uncertainty as to where the UK population effects would lie in the random effects distribution. Post-treatment transition probabilities used in the Markov model are presented in Appendix-5, Table A6.

Costs and utilities

Error! Reference source not found. summarises the cost per person for F2F (range £180-£907) and hybrid CBTs (range £205-£953) by various combinations of number and length of therapy sessions. For the same intensity, the treatment cost was higher for hybrid-CBT than F2F-CBT, given the additional multimedia components. The weighted mean cost for MM-CBT across the studies included in the systematic review was estimated to be £148/person (SD £141).

After the treatment period, for those who received CBT, the monthly cost of people in remission was estimated at £22 (SE £5), and that of people remained depressed was estimated at £79 (SE £20). For those who received TAU, monthly cost of people in remission was estimated at £41 (SE £12) and that of people remained depressed was estimated at £67 (SE £13) (Appendix-5, Table A6).

The estimated EQ-5D utility scores are listed in Table . The mean baseline utility score for people with depression was 0.65 (SD 0.23). At the end of treatment period, for people treated with any CBT option, the EQ-5D utility scores for those meeting the criteria for response and remission, response and no remission, no response and no remission were 0.92, 0.84 and 0.74, respectively. The corresponding utility scores for patients treated with TAU were 0.93, 0.80 and 0.74.

During the 5-years after the end of treatment, former patients in remission had a mean monthly QALY of 0.076 in the CBT groups and 0.077 in the TAU groups. For those who remained depressed, the mean monthly QALY for CBT and TAU groups were 0.064 and 0.058, respectively (Appendix-5, Table A6).

Cost-effectiveness analysis

Base-case analysis

Table 3 presents a summary of the base-case results of the incremental cost-effectiveness analysis based on the Markov Model over the 5-year time horizon. The second and third columns in Table 3 list the mean costs and QALYs of each intervention. Given the same treatment intensity, hybrid CBTs appeared to be more expensive and less effective compared with F2F-CBTs. Hence, all the hybrid CBTs are dominated by their F2F counterparts. Columns 4 lists ICERs of all the CBT options compared with TAU, while column 5 shows ICERs of some options compared with the next best option after excluding all the dominated (more costly but less effective than the comparator) and extended dominated (less cost-

effective than the next option) options. Appendix-5 Figure A2 illustrates the base-case cost-effectiveness plane. The CBT options on the cost-effectiveness frontier (the red line in Figure A2) are 6(sessions)×30(minutes) F2F, 8×30 F2F, 8×60 F2F and 16×60 F2F. The ICER of 6×30 F2F compared with TAU is £209/QALY. The estimated ICER for the comparison between 8×30 F2F and 6×30 F2F is £4,453/QALY. When the length of the 8-session F2F-CBT was extended from 30-minutes to 60-minutes, the ICER increased to £12,216/QALY. The16×60 F2F option was the most costly and most effective F2F intervention, resulting in an ICER of £43,072 per additional QALY for 16×60 F2F versus 8×60 F2F. Columns 6-7 in Table 3 indicate that 8×30 F2F has the highest net monetary benefit for thresholds of £20,000 and £30,000 per QALY.

PSA incorporated uncertainty in the parameter estimates to provide estimates of the probability that each intervention would be cost-effective at different acceptable ICER values. The CEAC in Figure 3. illustrates that for acceptable ICER values less than £25,290/QALY gained, the optimal option is 6×30 F2F. The probability of 6×30 F2F being the most cost-effective option declines as the acceptable value increases. At ICER values higher than £25,290/QALY, 16×60 F2F becomes the optimal option with the highest probability of being cost-effective. Figure 3 also shows that, using the NICE decision-making thresholds, 6×30 F2F has the highest probability (32.8%) of being cost-effective at £20,000/QALY, and 16×60 F2F has the highest probability (31.0%) of being cost-effective at £30,000/QALY. But the probability of these two CBTs being the most cost-effective option never exceeded 50%, indicating a high level of uncertainty.

Sensitivity analyses

We repeated the process for base-case analysis using the predicted effect mean for withdrawal rate. The overall results and findings (especially the cost-effectiveness of different CBT options) were comparable with the random effects mean model, and there were

relatively trivial changes to the values of the ICERs. Sensitivity analyses show the 6×30 F2F had the highest probability (32.5%) of being cost-effective at £20,000/QALY, and that 16×60 F2F had the highest probability (31.1%) of being cost-effective at £30,000/QALY. Appendix 6 lists the results of the sensitivity analysis where RR of mortality was assumed equal to one and the results remained robust.

Discussion

The base-case analysis found that the TAU arm yielded 3.501 QALYs over a five-year horizon at a cost of £3,325/person. When compared with TAU, F2F-CBT, hybrid-CBT and MM-CBT, all might be considered cost-effective given the NICE maximum acceptable ICER range of £20,000-£30,000 per QALY gained. This is consistent with findings of many recent systematic reviews for different forms of CBT [38,39]. Four F2F-CBT options –6×30, 8×30, 8×60 and 16×60 – were on the cost-effectiveness frontier, which describes the optimal pathway of moving from one strategy to a more costly one with highest increased outcome. 8×30 F2F generated the maximum net monetary benefits for thresholds of £20,000 and £30,000 per QALY.

However, after taking into account uncertainty surrounding the decision, the PSA illustrates that the least costly strategy (6×30 F2F) had the highest probability of being the most cost-effective strategy for lower thresholds (less than £25,290/QALY). As the threshold increases, though, the most effective strategy (16×60 F2F) becomes the preferred option. These results should be interpreted with caution, however, since none of the options reached a 50% probability of being most cost-effective, indicating a high level of uncertainty.

On the other hand, while we identified 65 F2F-CBT intervention arms with over 2,000 depression patients and 18 MM-CBT with 1,480 participants, there were only seven hybrid

CBTs arms identified in the systematic review. The sample size for hybrid-CBT was therefore much smaller than F2F and MM-CBT, adding another layer of uncertainty.

Akin to the dose-response relationship in pharmacological interventions, the intensity of CBT can influence treatment effectiveness; however, the optimal intensity of CBT is not clear [14,21]. In addition, treatment intensity also influences healthcare resource usage, thereby affecting the cost-effectiveness of CBT. In this study, a decision-analytic model was designed to compare the cost-effectiveness of F2F-CBT, hybrid-CBT and MM-CBT of various intensities with TAU. MM-CBT, although less intensive in resources demand, did not appear to be cost-effective in comparison with the alternatives. The optimal intensity of F2F-CBT appeared to be at the two ends of the spectrum considered. The alternatives in between, while incurred higher cost due to increased intensity, did not appear to produce sufficiently improved outcomes to justify the additional investment, although note that this finding is likely a result of the assumed linear model for intensity. To the best of our knowledge, this study was the first to provide a separate comparison of cost-effectiveness for different CBT delivery modes and intensities with TAU [39-41].

This study has several limitations. We used a loosely defined TAU as the common comparator in the analysis but reality is more complicated in that TAU is not a homogeneous service among different countries or over time. For example, the latest NICE Clinical Guidelines recommends both pharmacological and psychological treatments in usual practice [23]. This makes CBT part of the current TAU in the UK. Therefore, some of the earlier studies might be less comparable than later and especially future ones. Similarly, potential differences in clinical settings and effectiveness between countries were not explored in this analysis. In order to extract as much data as possible to estimate resource use, the analysis made use of all studies, regardless of country, and valued the resources using cost estimates

based on UK practices. However, some studies may have used treatment methods not typically used in the UK, which would bias towards unrealistically high costs.

We simplified the cost of CBT by using full-course costs for the completers and null costs for the non-completers. This may underestimate the cost of non-completers, as in practice, patients might withdraw in the middle of a course of treatment. On the other hand, this may also have a chance to overestimate the cost of completers, as completers who attended the final assessment may not attend all the sessions. Given the lack of empirical evidence of the actual number of sessions each group attended, we made the above assumption and assume a similar effect of withdrawal among all the intervention arms. We have also assumed that patients who do not achieve remission have the same long-term outcomes, regardless of whether they responded or not in the short-term. Relaxing this assumption is likely to improve the cost-effectiveness of interventions which had higher response rates, for a given remission rate.

Thirdly, the cost and utility scores for the post-treatment period were mainly derived from the CoBalT trial. Although this study provides the best available data for long-term follow-up after CBT and TAU treatment, the study only included patients with treatment-resistant depression who were taking antidepressant medication for at least 6 weeks. Our model adopted a broader population which includes adults with a primary diagnosis of depression and for whom CBT was considered by their GP, including these patients with treatment-resistant depression. Costs and utilities for the broader population might differ from those found in CoBalT, however because costs and utilities from the same population were used for all the intervention arms, we do not expect the qualitative findings to be sensitive to this, although the magnitude of differences may change. Further studies that measure costs and utilities with sufficiently long follow-up on less severely depressed populations are required for future models.

Within the hybrid CBTs, the materials used varied from books to smartphones, which is a source of heterogeneity. Their effects on reducing human involvement in therapy delivery could be very different. Moreover, for MM and hybrid CBTs, the analysis based on trials is unable to take into account the effect of potential improvement in accessibility and flexibility of the service. For example, it would prove cost-effective if hybrid-CBT could reach those who wouldn't have received treatment otherwise, or if the therapist time saved by one patient taking up MM-CBT could be used to treat another patient who is less amenable to multimedia.

Finally, we assumed a linear effect of intensity on the BDI scale. This was well estimated for F2F-CBT (based on 55 studies ranging in intensity from 1.6 to 19.8). However, we had insufficient evidence to estimate the effect of intensity for hybrid (5 studies ranging in intensity from 1.2 to 3.8), and instead assumed the effect of intensity was the same for hybrid and F2F. We may expect the effect of intensity to differ for hybrid and F2F due to the added multimedia components, in which case we could find that hybrid interventions were more cost-effective. Further studies exploring hybrid CBT at different intensities would be required to explore this further.

Financial constraints dictate that health care providers need to explore the integration of technology as part of services in order to increase access to, and effectiveness of, interventions such as CBT. Whilst the rapid development in technology has resulted in a proliferation of apps and online interventions for the self-management of chronic conditions including depression, few are evidence based. Moreover, previous work has shown that in order to increase engagement with, and maximise benefit from, computerised forms of CBT, human support is needed. By developing hybrid interventions that combine therapist input with use of technology to deliver tailored treatment, there is a potential for long-term gain. Future research that evaluates both clinical and cost-effectiveness of such approaches is

needed in order to inform provision and development of services. Meanwhile, the model constructed in this study can be easily adapted for use in future studies, to explore the long-term cost-effectiveness of other depression treatments.

References

1. Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* 2013;10(11):e1001547.
2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA.* 2003;289(23):3095-3105.
3. Ustun TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJ. Global burden of depressive disorders in the year 2000. *Br J Psychiatry.* 2004;184:386-392.
4. Ng CW, How CH, Ng YP. Depression in primary care: assessing suicide risk. *Singapore Med J.* 2017;58(2):72-77.
5. Angst J, Angst F, Stassen HH. Suicide risk in patients with major depressive disorder. *J Clin Psychiatry.* 1999;60 Suppl 2:57-62; discussion 75-56, 113-116.
6. Gaynes BN, West SL, Ford CA, et al. Screening for suicide risk in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2004;140(10):822-835.
7. Burcusa SL, Iacono WG. Risk for recurrence in depression. *Clinical Psychology Review.* 2007;27(8):959-985.
8. Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG. Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *Br J Psychiatry Suppl.* 1996(30):17-30.
9. McManus S, Bebbington P, Jenkins R, Brugha T. *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. .* Leeds: NHS Digital;2016.
10. Office for National statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland, mid-2017. 2018.
11. National Institute for Clinical Excellence (NICE). *Depression in adults: recognition and management 2009 [updated April 2016].* 2016.
12. McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. *PAYING THE PRICE, The cost of mental health care in England to 2026.* London: King's Fund;2008.
13. Health and Social Care Information Centre. Prescription Cost Analysis, England - 2017. 2018; file:///D:/pres-cost-anal-eng-2016-rep.pdf.
14. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can J Psychiatry.* 2013;58(7):376-385.
15. DeRubeis RJ, Gelfand LA, Tang TZ, Simons AD. Medications versus cognitive behavior therapy for severely depressed outpatients: Mega-analysis of four randomized comparisons. *American Journal of Psychiatry.* 1999;156(7):1007-1013.
16. Cuijpers P, Gentili C. Psychological treatments are as effective as pharmacotherapies in the treatment of adult depression: a summary from Randomized Clinical Trials and neuroscience evidence. *Research in Psychotherapy-Psychopathology Process and Outcome.* 2017;20(2):147-152.
17. Kaltenthaler E, Brazier J, De Nigris E, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technol Assess.* 2006;10(33).
18. Shapiro DA, Cavanagh K, Lomas H. Geographic Inequity in the Availability of Cognitive Behavioural Therapy in England and Wales. *Behavioural and Cognitive Psychotherapy.* 2003;31(2):185-192.
19. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine.* 2003;33(2):217-227.

20. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn Behav Ther.* 2009;38(4):196-205.
21. Cuijpers P, Huibers M, Ebert DD, Koole SL, Andersson G. How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders.* 2013;149(1-3):1-13.
22. Stulz N, Lutz W, Kopta SM, Minami T, Saunders SM. Dose-effect relationship in routine outpatient psychotherapy: does treatment duration matter? *J Couns Psychol.* 2013;60(4):593-600.
23. National Institute for Health and Clinical Excellence (NICE). Depression in adults: recognition and management (Clinical guideline [CG90]) updated 2016. 2009.
24. Lopez-Lopez JA, Davies SR, Caldwell DM, et al. The process and delivery of CBT for depression in adults: a systematic review and network meta-analysis. *Psychol Med.* 2019;49(12):1937-1947.
25. Wiles NJ, Thomas L, Turner N, et al. Long-term effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. *Lancet Psychiatry.* 2016;3(2):137-144.
26. Wiles N, Thomas L, Abel A, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *Lancet.* 2013;381(9864):375-384.
27. Kessler D, Lewis G, Kaur S, et al. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. *Lancet.* 2009;374(9690):628-634.
28. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF. Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica.* 2010;122(3):184-191.
29. Yiend J, Paykel E, Merritt R, Lester K, Doll H, Burns T. Long term outcome of primary care depression. *Journal of Affective Disorders.* 2009;118(1-3):79-86.
30. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. *Med Decis Making.* 2005;25(6):646-654.
31. Office for National Statistics. *Deaths registered in England and Wales: 2016.* 2017.
32. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord.* 2002;72(3):227-236.
33. NICE. National Institute for Health and Care Excellence: Guide to the methods of technology appraisal 2013. 2013.
34. Health Careers. Agenda for change - pay rates. 2017; https://www.healthcareers.nhs.uk/about/careers-nhs/nhs-pay-and-benefits/agenda-change-pay-rates. Accessed 21st August, 2017.
35. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Economics.* 2004;13(12):1203-1210.
36. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation.* USA: Oxford University Press; 2006.
37. Fenwick E, Marshall DA, Levy AR, Nichol G. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *Bmc Health Services Research.* 2006;6.
38. Paganini S, Teigelkotter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: A systematic review. *J Affect Disord.* 2018;225:733-755.
39. Brettschneider C, Djadran H, Harter M, Lowe B, Riedel-Heller S, Konig HH. Cost-utility analyses of cognitive-behavioural therapy of depression: a systematic review. *Psychother Psychosom.* 2015;84(1):6-21.

40. Barrett B, Byford S, Knapp M. Evidence of cost-effective treatments for depression: a systematic review. *J Affect Disord.* 2005;84(1):1-13.
41. Pirraglia PA, Rosen AB, Hermann RC, Olchanski NV, Neumann P. Cost-utility analysis studies of depression management: A systematic review. *American Journal of Psychiatry.* 2004;161(12):2155-2162.

Tables and figures

TABLE 1 TOTAL COST PER PERSON OF A FULL COURSE TREATMENT (2016/17 PRICE)

Number of sessions	F2F-CBT (30-min per session)	F2F-CBT (60-min per session)	hybrid-CBT (30-min per session)	hybrid-CBT (60-min per session)
6 sessions	£180	£358	£205	£389
8 sessions	£234	£468	£262	£501
10 sessions	£289	£577	£318	£614
12 sessions	£344	£687	£374	£727
14 sessions	£399	£797	£431	£840
16 sessions	£454	£907	£487	£953
MM-CBT	£148 (average cost per person)			

TABLE 2 UTILITY SCORES AT BASELINE AND AT THE END OF 4-MONTH TREATMENT PERIOD (SOURCE:[25,27])

EQ-5D utility scores	Mean	SD
Baseline	0.65	0.23
At the end of 4-month, with TAU treatment		
• Response and remission	0.93	0.09
• Response no remission	0.80	0.14
• No response no remission	0.74	0.19

At the end of 4-month, with CBT intervention		
• Response and remission	0.92	0.16
• Response no remission	0.84	0.12
• No response no remission	0.74	0.15

TABLE 3 ICERS AND NET MONETARY BENEFIT (5 YEARS AFTER INTERVENTION)

Interventions	**Cost**	**QALYs**	**ICERs compared to**		**Net monetary benefit**	
			Lowest cost (TAU)	**Next best option**	**£20,000 per QALY**	**£30,000 per QALY**
TAU	£3,325	3.501	-		£66,695	£101,705
F2F 6x30	**£3,400**	**3.859**	**£209/QALY**	**£209/QALY**	£73,780	£112,370
MM	£3,420	3.847	£275/QALY	Dominated	£73,520	£111,990
F2F 8x30	**£3,424**	**3.864**	**£273/QALY**	**£4,453/QALY**	£73,856	£112,496
Hybrid 6x30	£3,431	3.856	£299/QALY	Dominated	£73,689	£112,249
Hybrid 8x30	£3,457	3.861	£367/QALY	Dominated	£73,763	£112,373
F2F 10x30	£3,483	3.861	£439/QALY	Dominated	£73,737	£112,347
F2F 12x30	£3,500	3.868	£477/QALY	Extended Dominated	£73,860	£112,540
Hybrid 10x30	£3,518	3.858	£541/QALY	Dominated	£73,642	£112,222
F2F 6x60	£3,525	3.864	£551/QALY	Dominated	£73,755	£112,395
Hybrid 12x30	£3,536	3.865	£580/QALY	Dominated	£73,764	£112,414
Hybrid 6x60	£3,562	3.861	£658/QALY	Dominated	£73,658	£112,268
F2F 14x30	£3,566	3.862	£668/QALY	Dominated	£73,674	£112,294
F2F 8x60	**£3,572**	**3.875**	**£660/QALY**	**£12,216/QALY**	**£73,928**	**£112,678**
F2F 16x30	£3,576	3.872	£677/QALY	Dominated	£73,864	£112,584
Hybrid 14x30	£3,603	3.859	£777/QALY	Dominated	£73,577	£112,167
Hybrid 8x60	£3,612	3.872	£774/QALY	Dominated	£73,828	£112,548
Hybrid 16x30	£3,614	3.869	£785/QALY	Dominated	£73,766	£112,456
F2F 10x60	£3,699	3.866	£1,025/QALY	Dominated	£73,621	£112,281
F2F 12x60	£3,738	3.879	£1,093/QALY	Extended Dominated	£73,842	£112,632
Hybrid 10x60	£3,740	3.863	£1,146/QALY	Dominated	£73,520	£112,150
Hybrid 12x60	£3,783	3.876	£1,221/QALY	Dominated	£73,737	£112,497
F2F 14x60	£3,872	3.868	£1,490/QALY	Dominated	£73,488	£112,168
F2F 16x60	**£3,903**	**3.883**	**£1,513/QALY**	**£43,072/QALY**	£73,757	£112,587
Hybrid 14x60	£3,919	3.865	£1,632/QALY	Dominated	£73,381	£112,031

Hybrid 16x60	£3,953	3.880	£1,657/QALY	Dominated	£73,647	£112,447

Figure 1 Structure of the decision tree

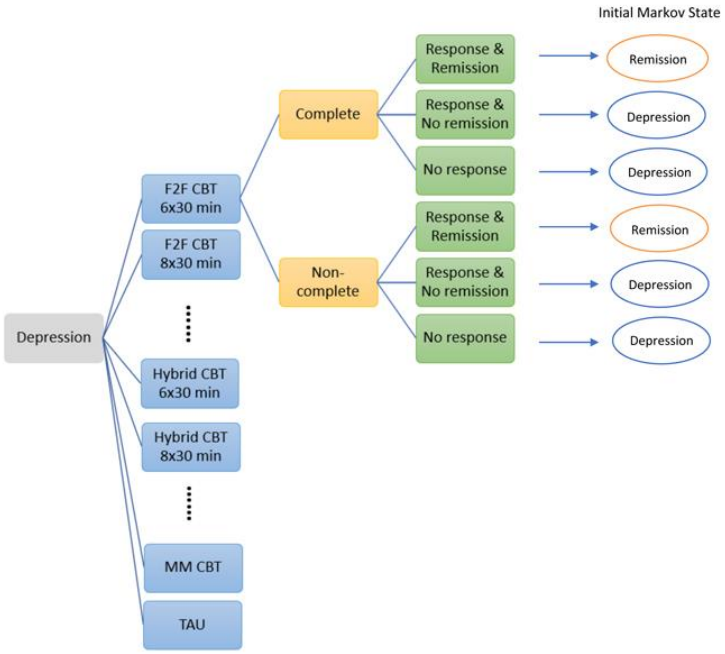


Figure 2 Structure of the Markov model

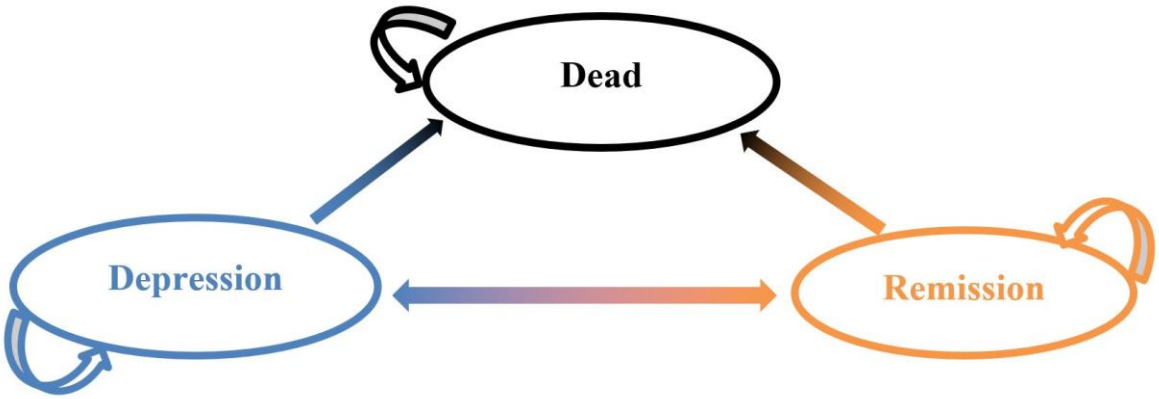
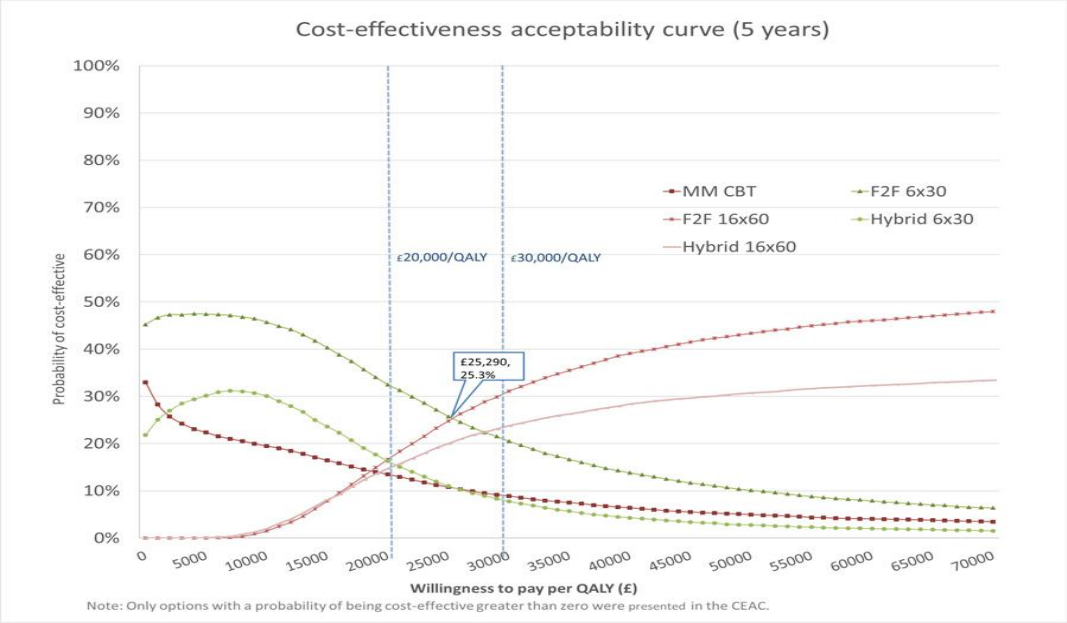


Figure 3 Cost-effectiveness acceptability curve (CEAC)


Cost-effectiveness acceptability curve (5 years)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Probability of cost-effective
£20,000/QALY
£30,000/QALY
£25,290, 25.3%
MM CBT
F2F 6x30
F2F 16x60
Hybrid 6x30
Hybrid 16x60
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
60000
65000
70000
Willingness to pay per QALY (£)
Note: Only options with a probability of being cost-effective greater than zero were presented in the CEAC.

Appendices

Appendix 1: Method for predicting Remission and Response Probabilities from Continuous Outcomes

The short-term economic model requires estimates of the probabilities of remission, response (without remission), and non-response, at end of treatment. Based on discussions with our project team we defined remission as a score of less than 10 points on the BDI. We define response as 50% reduction of BDI score at baseline. The RCTs identified in our systematic review primarily report continuous outcome measures of depression, on a variety of scales (BDI, HAMD, etc.), whereas reporting of response and remission outcomes was rare [1]. This makes the estimation of absolute probabilities of response, and remission for each of our decision options challenging. We chose to maximise our use of the available evidence by estimating response and remission probabilities from the outcomes on the BDI scale, the most commonly reported scale (see Appendix 3 for details). We used a bivariate Normal approximation to the baseline and follow-up score on the BDI for each intervention to obtain estimates of remission and response probabilities as follows.

Let X be the baseline score and Y the follow-up score. We define the change score $C = Y - X$. Response is defined as 50% reduction in score so

$$P(response) = P(C < -0.5X) = P(Y - X < -0.5X) = P(Y < 0.5X)$$

Remission is defined as a follow-up score less than a threshold value T (we assume $T = 10$), so

$$P(remission) = P(Y < T).$$

If we assume a multivariate Normal likelihood for the baseline and follow-up scores, then

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

where μ_X, μ_Y are the mean scores at baseline and follow-up respectively, σ_X, σ_Y are the standard deviation at baseline and follow-up respectively, and ρ is the correlation between baseline and follow-up measures.

Given an estimated mean change from baseline under TAU, η, and relative effect for intervention k compared with TAU, d_k, then the mean at follow-up is:

$$\mu_Y = \mu_X + \eta + d_k \qquad (1)$$

The probability of remission using a threshold T is therefore:

$$p(remission) = \Phi\left(\frac{T - (\mu_X + \eta + d_k)}{\sigma_Y}\right) \qquad (2)$$

Response is more complicated because it is a function of baseline score, and the follow-up score required to achieve response differs between individuals who differ in their baseline scores. We therefore condition on baseline score, X. Under a multivariate Normal assumption, the distribution for follow-up score, conditional on baseline score is

$$Y \mid X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1-\rho^2)\sigma_Y^2\right)$$

The probability of response conditional on baseline score is therefore

$$p(response \mid X = x) = \Pr(Y < 0.5x) = \Phi\left(\frac{0.5x - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)}{\sqrt{(1-\rho^2)}\sigma_Y}\right)$$

Averaging over the marginal distribution for baseline scores, $X \sim N\left(\mu_X, \sigma_X^2\right)$, gives:

$$p(response) = \underset{X}{E}\left[\Phi\left(\frac{0.5x - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)}{\sqrt{\left(1-\rho^2\right)}\sigma_Y}\right)\right]$$

Using the relationship: $\underset{X}{E}\left[\Phi\left(aX+b\right)\right] = \Phi\left(\frac{a\mu_X + b}{\sqrt{1+a^2\sigma_X^2}}\right)$ when X has a Normal distribution, it can be verified that

$$p(response) = \Phi\left(\frac{0.5x - \mu_Y\left(+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)}{\sqrt{\sigma_Y^2\left(1-\rho^2\right) + \sigma_X^2\left(0.5 - +\rho\frac{\sigma_Y}{\sigma_X}\right)^2}}\right)$$

Substituting equation (1) for μ_Y gives:

$$p(response) = \Phi\left(\frac{-\left(0.5\mu_X + \eta + d_k\right)}{\sqrt{\sigma_Y^2\left(1-\rho^2\right) + \sigma_X^2\left(0.5 - +\rho\frac{\sigma_Y}{\sigma_X}\right)^2}}\right) \qquad (3)$$

For TAU baseline and follow-up scores on BDI were estimated from an RCT chosen to be representative of UK practise (see Appendix 2 for details). For all other interventions relative effects on mean change in BDI scores compared with TAU were obtained from a re-analysis of the data from the systematic review described in Appendix 3 1. These relative effects were applied to the follow-up BDI score for TAU to obtain absolute BDI follow-up scores for all interventions. Equations (2) and (3) were then applied to obtain the required transition

probabilities for each intervention. Simulation was used to capture the uncertainty in the estimated mean BDI scores which was then propagated through to the transition probabilities, and it was these simulated values that were used in the probabilistic sensitivity analysis.

Appendix 2: Remission and response probabilities for TAU

We used results from the IPCRESS RCT to inform TAU in the short-term treatment model, as representative of a UK population relevant to our decision population.[2] The control arm is described as wait-list, however our collaborators felt that it in fact represented current UK practice, and would be better described as treatment as usual (TAU). Furthermore, we had access to the IPD from IPCRESS, which allowed us to check our approximation method.

Table A1 shows mean and standard deviation for BDI at baseline in each arm, and across arms. Baseline BDI is balanced across arms, and we therefore assume a mean BDI of 33.175 with SD=8.805 in our TAU reference population. In our network meta-analysis we assumed a correlation between baseline and follow-up scores of 0.65, based on studies in our systematic review that provided summaries at baseline, follow-up, and also gave change from baseline summaries (and hence allowed a correlation to be estimated) [1]. Table A1 shows that in IPCRESS the correlation between baseline and 4-month (end of treatment) follow-up BDI scores was lower than this in both the TAU and CBT arms (with correlations of 0.56 and 0.40 resp.). We chose to use the higher value of 0.65 to be consistent with the network meta-analysis. The mean change from baseline on TAU was -11.27 with standard error 1.16. We assume the SD for BDI scores at follow-up is 12.859, based on the SD for BDI across both arms of IPCRESS at end of treatment (4-months follow-up).

TABLE A1: SUMMARY STATISTICS FROM THE IPCRESS RCT AT BASELINE, AND END OF TREATMENT (4 MONTHS FOLLOW-UP).

	Baseline			**End of treatment (4 months Follow-up)**		**Change from baseline**	
	mean BDI	SD BDI	Correlation baseline and 4month BDI scores	mean BDI	SD BDI	mean change from baseline	SE (change from baseline)
TAU arm	33.527	9.270	0.562	22.253	13.466	-11.274	1.16
CBT arm	32.826	8.336	0.401	14.509	11.204	-21.92	0.973
Pooled	33.175	8.805		18.063	12.859		

Based on the figures reported in IPCRESS we can estimate the remission and response probabilities under a Normal approximation (Appendix 1), giving the results shown in Table A2, for various assumed values for the correlation between baseline and 4-month follow-up. In IPCRESS, remission probabilities were higher than response probabilities, and both response and remission probabilities were higher on the CBT arm compared with the TAU arm Table A2. The Normal approximation slightly under-estimates the probability of remission, although the estimates are fairly close and well within the confidence limits. Similarly, the Normal approximation underestimates the response probabilities, most notably for the TAU arm, but the approximation is still within the confidence bounds. Although the

approximation is not perfect, we considered it to be sufficiently accurate to use it to obtain estimates of response and remission probabilities based on population characteristics from IPCRESS (see Table A2 legend) and the estimates of mean difference in change from baseline on the BDI scale, obtained from our network meta-analysis (Appendix 3).

TABLE A2: COMPARISON OF RESPONSE AND REMISSION PROBABILITIES ESTIMATED DIRECTLY FROM IPCRESS DATA, AND THOSE OBTAINED BY THE NORMAL APPROXIMATION METHOD ASSUMING: BASELINE MEAN BDI = 33.175 WITH SD=8.805, MEAN CHANGE FROM BASELINE AS FOUND IN IPCRESS, AND FOLLOW-UP SD IN BDI OF 12.859. REMISSION IS ASSUMED TO BE DEFINED AS BDI <10. APPROXIMATIONS TO RESPONSE PROBABILITIES ARE PRESENTED FOR CORRELATIONS BETWEEN BASELINE AND FOLLOW-UP MEASURES OF 0.5, 0.6, AND 0.65.

	IPCRESS	**Normal approximation to p(response)**			**IPCRESS**	**Normal approximation to p(remission)**
	p(response) (95%CI)	**Cor=0.5**	**Cor=0.6**	**Cor=0.65**	**p(remission) (95% CI)**	
TAU arm	0.37 (0.27, 0.47)	0.3216	0.3137	0.3093	0.21 (0.13, 0.30)	0.181
CBT arm	0.60 (0.51, 0.69)	0.5770	0.5809	0.5832	0.38 (0.29, 0.46)	0.344

Appendix 3: Relative effects on the BDI scale

The RCTs identified in the systematic review form a network of intervention comparisons, including CBT interventions, TAU, wait list, no treatment and placebo [1]. We conducted a

network meta-analysis (NMA) which is a method to simultaneously estimate intervention effects respecting the randomisation in the RCTs [3]. Lopez-Lopez report results from an NMA on the standardised mean difference (SMD) scale because different measures of depression are used in different studies [1]. The economic model however is based on the BDI depression scale, and so we require estimates of intervention effects on the BDI scale.

Not all studies report the BDI outcome, however all studies report at least one depression measure, and many studies report multiple outcomes. Figure A1 shows a network of outcomes plot, which indicates which outcomes are measured together in the same study (indicated with a connecting line). This gives us the opportunity to estimate mapping coefficients between intervention effects on different depression scales, so that all studies can be combined in a single analysis to estimate intervention effects on the BDI scale (regardless as to whether they included BDI as an outcome) [4]. Lopez-Lopez [1] found some weak evidence on the SMD scale that intensity of intervention may be an effect modifier for F2F CBT but not for MM CBT, where intensity is defined as:

$$intensity = no.sessions \times length\ of\ \ sessions(mins)/100$$

We therefore explored the inclusion of intensity as a covariate in the NMA model on the BDI scale. Due to limited data, we were unable to assess whether intensity was associated with intervention effect for Hybrid CBT. However, since Hybrid CBT involved therapist contact, after discussion with our project team we made the assumption that the interaction with intensity for Hybrid was equal to that for F2F CBT (but the main intervention effects were assumed to differ between F2F and Hybrid CBT). Table A3 presents parameter estimates from this model (on the BDI scale) along with estimates of the mapping coefficients. We found that results from the mapping approach were similar to those obtained using a standardised mean difference scale, but had the advantage that the resulting estimates were on an interpretable depression scale where thresholds for remission were available. The relative

effects from Table A3 were combined with information on TAU (see Appendix 2) to obtain a distribution of absolute scores on the BDI for each intervention at follow-up for a given baseline distribution of BDI scores. These were then used to construct predicted probabilities of response and remission required in the economic model based on a bivariate Normal assumption (equations (2) and (3) in Appendix 1). The resulting transition probabilities are summarised in Table A4 which shows the mean and 95% CrIs from the PSA samples.

Table A3: Posterior summaries (posterior mean and 95% credible interval) for mean difference in change from baseline to end of treatment on BDI scale, and estimated mappings between intervention effects on the different depression scales. Abbreviations: BDI: Beck Depression Inventory (any version); BSI: Brief Symptom Inventory; CES-D: Center for Epidemiologic Studies Depression Scale (CES-D); HAM-D: Hamilton Depression Scale; MADRS: Montgomery Asberg Depression Rating Scale; MMPI-depression: Minnesota Multiphasic Personality Inventory (depression subscale); PHQ-8 and PHQ-9: Patient's Health Questionnaire (8-item and 9-item versions, respectively).

Posterior mean difference in change from baseline to end of treatment on BDI (95% CrI)	**Main Effects for F2F, Hybrid, MM** **Interaction with intensity (F2F, Hybrid)**
F2F CBT	-5.535 (-10.42, -0.707)
Hybrid	-5.018 (-10.81, 0.608)
MM	-4.304 (-9.565, 0.835)
Interaction with intensity (F2F and Hybrid)	-0.496 (-0.922, -0.063)
Between studies sd	5.70 (4.61, 7.07)
No. studies for NMA with interaction with intensity	54
Mappings	
Mapping BSI->BDI	0.902 (0.684,1.141)

Mapping CES-D->BDI	1.053 (0.955,1.163)
Mapping HAM-D->BDI	0.926 (0.5451,1.618)
Mapping MADRS->BDI	2.564 (0.159,15.75)
Mapping MMPI->BDI	0.223 (0.076,0.548)
Mapping PHQ8->BDI	1.169 (0.800,1.57)
Mapping PHQ9->BDI	0.902 (0.684,1.141)

FIGURE A1: A NETWORK PLOT SHOWING WHERE OUTCOMES SCALES HAVE BEEN MEASURED IN THE SAME RCT (INDICATED BY LINES). BASED ON 70 STUDIES. ABBREVIATIONS: BDI: BECK DEPRESSION INVENTORY (ANY VERSION); BSI: BRIEF SYMPTOM INVENTORY; CES-D: CENTER FOR EPIDEMIOLOGIC STUDIES

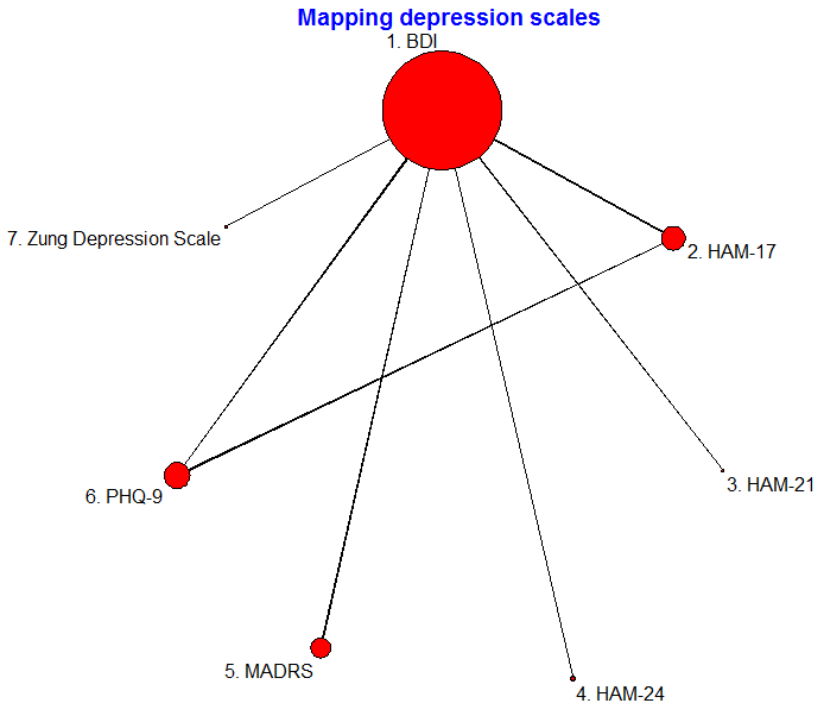


TABLE A4: ESTIMATED RESPONSE AND REMISSION PROBABILITIES AT THE END OF TREATMENT

Interventions	**Probability of response & remission (95% CI)**	**Probability of response but no remission (95% CI)**	**Probability of no response & no remission (95% CI)**
TAU	0.18 (0.14,0.23)	0.13 (0.10, 0.16)	0.69 (0.60, 0.76)
F2F CBT 6 sessions 30mins	0.3438 (0.2154,0.4891)	0.2048 (0.153,0.2335)	0.4514 (0.2771,0.6316)
F2F CBT 8 sessions 30mins	0.3521 (0.2258,0.4937)	0.2075 (0.1589,0.2336)	0.4405 (0.2724,0.6153)
F2F CBT 10 sessions 30mins	0.3604 (0.2362,0.4988)	0.21 (0.1645,0.2337)	0.4295 (0.2673,0.5993)
F2F CBT 12 sessions 30mins	0.3689 (0.2463,0.5044)	0.2125 (0.1697,0.2338)	0.4187 (0.2616,0.584)
F2F CBT 14 sessions 30mins	0.3774 (0.2562,0.5105)	0.2148 (0.1747,0.2338)	0.4078 (0.2557,0.5691)
F2F CBT 16 sessions 30mins	0.386 (0.2658,0.5165)	0.2169 (0.1793,0.2339)	0.397 (0.2498,0.5549)
F2F CBT 6 sessions 60mins	0.3689 (0.2463,0.5044)	0.2125 (0.1697,0.2338)	0.4187 (0.2616,0.584)
F2F CBT 8 sessions 60mins	0.386 (0.2658,0.5165)	0.2169 (0.1793,0.2339)	0.397 (0.2498,0.5549)
F2F CBT 10 sessions 60mins	0.4035 (0.2854,0.5297)	0.2208 (0.188,0.2339)	0.3757 (0.2371,0.5266)
F2F CBT 12 sessions 60mins	0.4212 (0.3038,0.5453)	0.224 (0.1956,0.234)	0.3548 (0.2225,0.5006)
F2F CBT 14 sessions 60mins	0.4391 (0.3207,0.5624)	0.2265 (0.2019,0.234)	0.3344 (0.2072,0.4774)
F2F CBT 16 sessions 60mins	0.4572 (0.3364,0.5813)	0.2283 (0.2072,0.234)	0.3145 (0.1907,0.4563)
Hybrid CBT 6 sessions 30mins	0.3308 (0.182,0.5082)	0.1984 (0.1327,0.2336)	0.4708 (0.2579,0.6853)
Hybrid CBT 8 sessions 30mins	0.3389 (0.1889,0.5166)	0.201 (0.1371,0.2337)	0.46 (0.2497,0.674)

Hybrid CBT 10 sessions 30mins	0.3472 (0.1958,0.5251)	0.2036 (0.1414,0.2338)	0.4493 (0.2415,0.6628)
Hybrid CBT 12 sessions 30mins	0.3555 (0.2023,0.5334)	0.2059 (0.1453,0.2338)	0.4385 (0.2336,0.6523)
Hybrid CBT 14 sessions 30mins	0.3639 (0.2089,0.5425)	0.2082 (0.1493,0.2339)	0.4279 (0.2251,0.6418)
Hybrid CBT 16 sessions 30mins	0.3724 (0.2155,0.5524)	0.2103 (0.153,0.2339)	0.4172 (0.2161,0.6315)
Hybrid CBT 6 sessions 60mins	0.3555 (0.2023,0.5334)	0.2059 (0.1453,0.2338)	0.4385 (0.2336,0.6523)
Hybrid CBT 8 sessions 60mins	0.3724 (0.2155,0.5524)	0.2103 (0.153,0.2339)	0.4172 (0.2161,0.6315)
Hybrid CBT 10 sessions 60mins	0.3897 (0.2281,0.5719)	0.2141 (0.1601,0.2339)	0.3963 (0.1988,0.6118)
Hybrid CBT 12 sessions 60mins	0.4071 (0.24,0.5921)	0.2172 (0.1665,0.2339)	0.3757 (0.1817,0.5935)
Hybrid CBT 14 sessions 60mins	0.4247 (0.251,0.6139)	0.2197 (0.172,0.234)	0.3556 (0.1641,0.577)
Hybrid CBT 16 sessions 60mins	0.4425 (0.2613,0.6364)	0.2215 (0.1765,0.234)	0.336 (0.1468,0.5616)
MM CBT	0.29 (0.16, 0.45)	0.18 (0.12, 0.23)	0.53 (0.32, 0.73)

Appendix 4: Treatment costs

TABLE A5: UNIT COST OF OTHER MATERIALS USED IN THE PROCESS OF DELIVERY OF CBTS (2016/17 PRICE)

Item/Service	Unit cost	Sources
Book	£15 / book	Estimated from market price of similar products
CD	£10 / CD	
DVD	£20 / DVD	
Online CBT/Smartphone app	£50 / subscription	License fee of Beating the Blues [5]
Telephone	£0.12 / minute	BT market price personal landline cost [6]
Videoconference	£0.18 / minute	BT market price of data and Skype data usage per minute [6,7]

Appendix 5: Markov Model Parameters and results

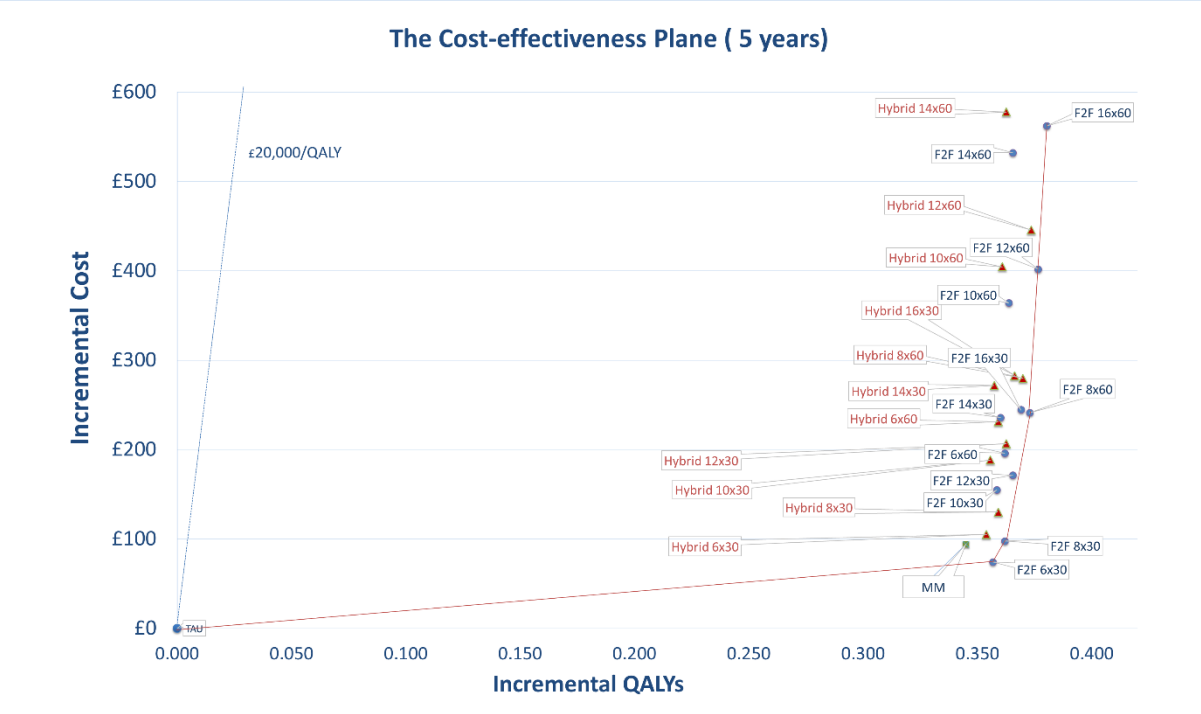
TABLE A6: MARKOV MODEL PARAMETERS (AFTER THE END OF TREATMENT)

Probability of remission (monthly)	Value (95% CI)	Source
Remission within 6 months after the end of treatment (TAU)	2.03% (1.22%, 3.15%)	[8]
Relative risk of any CBT compared with TAU (within 6 months)	2.17 (1.30, 3.63)	[8]
Remission from month 7 to 5-year after the end of treatment (TAU)	0.38% (0.19%, 0.69%)	[8]
Relative risk of any CBT compared with TAU (7-month to 5-year)	1.42 (0.68, 2.98)	[8]
Probability of relapse (monthly)		[8]
Relapse within 6 months (TAU)	8.90% (4.47%, 15.72%)	[8]
Relative risk of any CBT compared with TAU (within 6 months)	0.53 (0.27, 1.02)	[8]

Relapse from month 7 to 5-year after the end of treatment (TAU)	2.54% (1.20%, 4.70%)	[8]
Relative risk of any CBT compared with TAU (7-month to 5-year)	0.96 (0.61, 1.50)	[8]
Mortality (monthly)		[8]
Male (Age-standardised mortality rates 2016)	0.09%	[9]
Female (Age-standardised mortality rates 2016)	0.07%	[9]
Female percentage	65.10% (58.90%, 70.94%)	[10]
Relative risk of depression compared with remission	1.52 (1.45, 1.59)	[11]
Monthly age-standardised mortality (Remission)	0.08%	
Monthly age-standardised mortality (Depression)	0.12%	
Costs		
Monthly cost (Depression in TAU group)	£66.73 (SE £13.07)	[8]
Monthly cost (Remission in TAU group)	£41.32 (SE £12.29)	[8]
Monthly cost (Depression in any CBT group)	£78.88 (SE £20.12)	[8]
Monthly cost (Remission in any CBT group)	£21.77 (SE £5.25)	[8]
QALY		
Monthly QALY gained (Depression in TAU group)	0.058 (SE 0.002)	
Monthly QALY gained (Remission in TAU group)	0.077 (SE 0.001)	[8,12]
Monthly QALY gained (Depression in any CBT group)	0.064 (SE 0.002)	[8,12]
Monthly QALY gained (Remission in any CBT group)	0.076 (SE 0.002)	[8,12]
Yearly discount rate		
Cost discount rate	3.5%	[13]

QALY discount rate	3.5%	13
Willingness-to-pay thresholds	£20,000-£30,000/QALY	13

Figure A2 The Cost-effectiveness Plane with TAU as comparator (5 years)



Appendix 6: Results of sensitivity analysis

TABLE A7: RESULTS OF SENSITIVITY ANALYSIS USING RR OF MORTALITY =1, ICERS AND NET MONETARY BENEFIT (5 YEARS AFTER INTERVENTION)

Interventions	**Cost**	**QALYs**	**ICERs compared to**		**Net monetary benefit**	
			Lowest cost (TAU)	**Next best option**	**£20,000 per QALY**	**£30,000 per QALY**
TAU	£3,360	3.533	-		£67,299	£102,629
F2F 6x30	**£3,431**	**3.886**	**£201/QALY**	**£201/QALY**	£74,284	£113,142
MM	£3,451	3.875	£269/QALY	Dominated	£74,045	£112,794
F2F 8x30	**£3,454**	**3.891**	**£263/QALY**	**£4,713/QALY**	£74,359	£113,266
Hybrid 6x30	£3,462	3.883	£292/QALY	Dominated	£74,202	£113,034
Hybrid 8x30	£3,486	3.888	£357/QALY	Dominated	£74,273	£113,153
F2F 10x30	£3,511	3.887	£428/QALY	Dominated	£74,236	£113,110

F2F 12x30	£3,527	3.894	£464/QALY	Extended Dominated	£74,351	£113,290
Hybrid 10x30	£3,545	3.885	£527/QALY	Dominated	£74,150	£112,997
F2F 6x60	£3,552	3.891	£538/QALY	Dominated	£74,261	£113,167
Hybrid 12x30	£3,562	3.891	£566/QALY	Dominated	£74,262	£113,175
Hybrid 6x60	£3,588	3.888	£642/QALY	Dominated	£74,172	£113,051
F2F 14x30	£3,592	3.889	£653/QALY	Dominated	£74,187	£113,077
F2F 8x60	**£3,597**	**3.901**	**£644/QALY**	**£14,166/QALY**	**£74,418**	**£113,425**
F2F 16x30	£3,600	3.897	£660/QALY	Dominated	£74,346	£113,319
Hybrid 14x30	£3,628	3.886	£760/QALY	Dominated	£74,100	£112,964
Hybrid 8x60	£3,635	3.898	£755/QALY	Dominated	£74,324	£113,303
Hybrid 16x30	£3,638	3.895	£770/QALY	Dominated	£74,254	£113,200
F2F 10x60	£3,720	3.892	£1003/QALY	Dominated	£74,125	£113,048
F2F 12x60	£3,757	3.904	£1070/QALY	Extended Dominated	£74,327	£113,368
Hybrid 10x60	£3,761	3.890	£1124/QALY	Dominated	£74,032	£112,928
Hybrid 12x60	£3,801	3.901	£1197/QALY	Dominated	£74,226	£113,239
F2F 14x60	£3,888	3.894	£1463/QALY	Dominated	£73,991	£112,930
F2F 16x60	**£3,917**	**3.908**	**£1487/QALY**	**£46,396/QALY**	£74,236	£113,312
Hybrid 14x60	£3,934	3.891	£1602/QALY	Dominated	£73,891	£112,804
Hybrid 16x60	£3,966	3.905	£1631/QALY	Dominated	£74,129	£113,176

References:

1. Lopez-Lopez JA, Davies SR, Caldwell DM, et al. The process and delivery of CBT for depression in adults: a systematic review and network meta-analysis. *Psychol Med.* 2019;49(12):1937-1947.
2. Kessler D, Lewis G, Kaur S, et al. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. *Lancet.* 2009;374(9690):628-634.
3. Caldwell DM, Welton NJ. Approaches for synthesising complex mental health interventions in meta-analysis. *Evid Based Ment Health.* 2016;19(1):16-21.
4. Ades AE, Lu GB, Dias S, Mayo-Wilson E, Kounali D. Simultaneous synthesis of treatment effects and mapping to a common scale: an alternative to standardisation. *Research Synthesis Methods.* 2015;6(1):96-107.
5. Littlewood E, Duarte A, Hewitt C, et al. A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. *Health Technology Assessment.* 2015;19(101):1-+.
6. British Telecommunications. BT Consumer Price Guide. 2018; Access: https://img01.products.bt.co.uk/content/dam/bt/storefront/pdfs/bt-consumer-price-guide-15.pdf archived at http://www.webcitation.org/719ZcyG27.
7. Skype. How much bandwidth do I need for calls? 2018; access: https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need?q=how+much+data archived at http://www.webcitation.org/719ZKRC2t

8. Wiles NJ, Thomas L, Turner N, et al. Long-term effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. *Lancet Psychiatry.* 2016;3(2):137-144.
9. Office for National Statistics. *Deaths registered in England and Wales: 2016.* 2017.
10. McManus S, Bebbington P, Jenkins R, Brugha T. *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014.* . Leeds: NHS Digital;2016.
11. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord.* 2002;72(3):227-236.
12. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. *Br J Psychiatry.* 2010;197(4):297-304.
13. NICE. National Institute for Health and Care Excellence: Guide to the methods of technology appraisal 2013. 2013.