Cost-effectiveness of structured group psychoeducation versus unstructured group support for bipolar disorder: results from a multi-centre pragmatic randomised controlled trial.

EM Camacho[1], D Ntais[1], S Jones[2], L Riste[3], R Morriss[45], F Lobban[2], LM Davies[1]

[1] Manchester Centre for Health Economics, The University of Manchester, UK
[2]Spectrum Centre for Mental Health Research, Lancaster University, UK
[3]School of Psychological Sciences, The University of Manchester, UK
[4]Institute of Mental Health, University of Nottingham, UK
[5]Nottinghamshire Healthcare NHS Trust, UK

Corresponding author: Elizabeth Camacho, Manchester Centre for Health Economics, Division of Population Health, Health Services Research, and Primary Care, University of Manchester, Oxford Road, Manchester, UK, M13 9PL. elizabeth.camacho@manchester.ac.uk

Funding: This paper presents independent research funded as part of the PARADES Programme Grant for Applied Research (RP-PG-0407-10389) by the National Institute for Health Research (NIHR), Department of Health, England. It received further support from primary care trusts, mental health trusts, the Mental Health Research Network and Comprehensive Local Research Networks in the East Midlands and North West England. The views expressed by the authors do not necessarily reflect those of the National Institute for Health Research, the National Health Service nor the Department of Health in England.RM was partly funded by the NIHR Collaboration for Leadership in Applied Health Research and Care East Midlands. We received further support from primary care trusts, mental health trusts, the Mental Health Research Network, and Comprehensive Local Research Networks in the East Midlands and North West England. The views expressed by the authors do not necessarily reflect those of the NIHR, the National Health Service (NHS), or the Department of Health in England.

ABSTRACT

Background Bipolar disorder (BD) costs the English economy an estimated £5.2billion/year, largely through incomplete recovery. This analysis estimated the cost-effectiveness of group psychoeducation (PEd), versus group peer support (PS), for treating BD.

Methods A 96-week pragmatic randomised controlled trial (RCT), conducted in NHS primary care. The primary analysis compared PEd with PS, using multiple imputed datasets for missing values. An economic model was used to compare PEd with treatment as usual (TAU). The perspective was Health and Personal Social Services.

Results Participants receiving PEd (n=153) used more (costly) health-related resources than PS (n=151) (net cost per person £1098 (95% CI, £252-£1943)), with a quality-adjusted life year (QALY) gain of 0.023 (95% CI, 0.001-0.056). The cost per QALY gained was £47,739. PEd may be cost-effective (versus PS) if decision makers are willing to pay at least £37,500 per QALY gained. PEd costs £10,765 more than PS to avoid one relapse. The economic model indicates that PEd may be cost-effective versus TAU if it reduces the probability of relapse (by 15%) or reduces the probability of *and* increases time to relapse (by 10%).

Limitations Participants were generally inconsistent in attending treatment sessions and low numbers had complete cost/QALY data. Factors contributing to pervasive uncertainty of the results are discussed.

Conclusions This is the first economic evaluation of PEd versus PS in a pragmatic trial. PEd is associated with a modest improvement in health status and higher costs than PS. There is a high level of uncertainty in the data and results.

Keywords: cost-effectiveness; bipolar disorder; psychoeducation; peer support

Highlights:

- The cost-effectiveness of group psychoeducation (PEd) to treat BD is unknown
- Data were collected as part of a pragmatic randomised controlled trial (RCT)
- The control treatment in the RCT was group peer support (PS)
- PEd may be cost-effective but this is uncertain
- Further economic data may help to address this uncertainty

Conflicts of interest: All authors declare that they have no conflicts of interest.

1. INTRODUCTION

Bipolar disorder (BD) is the 18th leading cause of disability (years lived with) for any health problem (1). BD has been estimated to cost the English economy £5.2 billion annually, largely due to incomplete recovery as a result of inadequate treatment(2).

Traditionally, medications such as lithium carbonate have been used to prevent episodes of illness and are still recommended for this purpose(3) but are only partially effective. In high income countries people with BD often have the opportunity to access peer support (PS) groups, in addition to medication and support from health professionals. People with BD value such support (4) which may improve self-efficacy and be effective in managing many long term health conditions (5). Clinical guidelines for the management of BD recommend psychological treatment for the prevention of relapse in addition to pharmacotherapy (6;7). Current guidelines for England recommend manualised evidence-based psychological interventions developed specifically for BD as a component of long-term management of BD (3).

Because relatively large groups (10 to 18 people) can undertake treatment together, manualised psychological interventions delivered to groups, for example group psychoeducation (PEd), may be an efficient option for mental health services to improve outcomes for people with BD. Although therapists need to be trained to run the groups, the intervention is highly manualised and training is less intensive than for other psychological approaches such as individual cognitive behavioural therapy (CBT)(8). The existing evidence base for group psychological treatments in preventing relapses for BD is inconsistent. The first trial of PEd in a group of people with BD, showed clinical and cost-effectiveness for all types of bipolar relapse(9-11). Generally however findings from previous trials of PEd have been heterogeneous in design, often involving a small number of participants and short follow-up periods. They reported inconsistent effects on mood symptoms, quality of life, or functioning (12).

The analyses described here were part of an integrated clinical and economic randomised controlled trial (RCT) of the effectiveness of PEd compared to PS (trial acronym: PARADES) (13). Results suggested that while PEd was no more clinically effective than PS it was more acceptable to participants. PS, rather than treatment as usual (TAU), was chosen as the comparator in the RCT as the clinical aim was to compare PEd with an unstructured group-based intervention, matching for attentional effects. However, PS is not standard care in the UK, opportunities to access PS are highly variable, and to our knowledge there is no plan to establish PS in clinical practice. Furthermore a key problem identified in the National Institute for Health and Care Excellence (NICE) guideline for bipolar disorder (CG185) was that there was insufficient evidence to model the relative cost-effectiveness of psychological therapies compared to TAU (3). This is important because the decision problem facing providers is whether or not to provide PEd *in addition to* TAU, rather than which intervention to provide between PEd and PS. In order to address this, a simple economic model and threshold analysis were used to synthesise results of the RCT with clinical literature. This approach provides additional information for settings or patient groups where access to psychological therapies such as that provided by PS is limited (13).

1.1 Aims and objectives The overall aim of the economic evaluation was to explore the likelihood that PEd is cost-effective. Specific objectives were to:

- Estimate and compare the costs and quality-adjusted life years (QALYs) for participants in the intervention (PEd) and control (PS) groups at baseline and follow-up;
- Use RCT data to assess whether there were differences in the relative cost-effectiveness of PEd compared with PS;
- As PS is unlikely to be adopted as routine practice, explore whether PEd could be cost-effective compared with TAU, using a probabilistic simulation (economic) model.

2. METHODS

The design and results of the PARADES RCT have been described in detail elsewhere (14;15). Key features of the trial are described in brief here.

2.1 Population The target population was people with bipolar 1 or 2 affective disorder, not currently in episode but at increased risk of further relapse (defined as having had at least one episode in the last 24 months). The mean age of the sample was 45 years and 42% were male. Detailed characteristics of the trial sample are reported alongside the clinical results of the trial (14). The recruitment strategy was deliberately broad to ensure that the sample reflected a diversity of people with BD.

2.2 Intervention A detailed description of the interventions has been published as part of the clinical evaluation(14). A dual-region, single-blind, RCT compared the effectiveness of:

- 21 weekly bipolar group PEd sessions delivered by two health professionals (nurse, psychiatrist, psychologist or occupational therapist) and a service-user (SU) facilitator, plus TAU.
- 21 weekly unstructured bipolar group PS sessions delivered by two health professionals and a 'service-user' facilitator, plus TAU(15).

Both treatments were delivered over a period of up to 26 weeks (allowing for holiday periods) (15). As part of PEd, a manual covering the 21 sessions was produced, with a session-specific handout for each session and opportunities for further 'homework tasks'. For PS, a short manual outlining group rules was used and participants collectively decided upon an agenda for discussion at each session. Both interventions were delivered adjunctively to treatment prescribed by their usual physician. Consecutively eligible people were individually randomised to either treatment, with minimisation by number of previous bipolar episodes (1-7, 8-19, 20+), and stratification by clinical site. Participants in both arms were followed-up over 96 weeks to allow sufficient time to observe changes in primary and secondary outcomes, while being feasible to implement.

2.3 Outcomes The primary measure of health benefit was the quality-adjusted life year (QALY). This was calculated based on participants' responses on the EQ-5D (3-level version) questionnaire and the published tariff of associated utility values for the UK (Measurement and Valuation of Health A1 tariff) (16;17). Participants were asked to complete the EQ-5D at the scheduled baseline, 32, 64 and 96 weeks assessments. QALYs were estimated in line with NICE recommendations for economics evaluations (18).

The EQ-5D has been shown to correlate with primary clinical outcome measures in previous trials of psychosis(19;20) and BD(21). Furthermore baseline utility values derived from the EQ-5D for the RCT sample were well correlated with clinical outcome measures (HADS-anxiety: Pearson coefficient= -0.474, p=0.000; HADS-depression: Pearson coefficient= -0.481, p=0.000; n=245; whether the participant relapsed: Pearson coefficient =-0.234, P = 0.048).

A secondary measure of health benefit was time to next bipolar episode, measured using 16-weekly Structured Clinical Interview for DSM-IV (SCID)-Longitudinal Interval Follow-up Evaluation (LIFE) interviews (22) to generate scores of mania and depression (23;24), or clinical notes where available.

2.4 Measuring costs The costs of the PEd and PS interventions included the costs of training staff and service-users to deliver the intervention (trainer and trainee time), the costs of delivering the intervention (time of staff and service users (SUs) to run the group sessions, plus materials), and the costs of supervision (time of supervisors and supervisees). The resources required to deliver the interventions were calculated based on their description as per the trial protocol (15). The amount of time dedicated to training and supervisions were recorded by the individuals who delivered the training and supervisions. The costs per unit of time were estimated by role (e.g. Psychiatrist, Therapist) from published NHS reference costs (25). The cost per group was calculated by dividing the total cost by the number of groups run. The cost per participant was calculated by dividing the cost per group by the actual number of participants allocated to the group. This approach assumes that there are no savings associated with participants who did not attend one or more of the planned group sessions. The difference in cost between the PEd and PS interventions equated to the cost of reproducing the PEd manual given to participants.

Costs associated with healthcare resource use were also considered in this analysis. These included the costs of the following: inpatient psychiatric and non-psychiatric care (including intensive care, emergency and crisis admissions); hospital outpatient and day hospital attendances; primary care contacts with the GP and GP practice staff; prescription medicines; community mental health care contacts; social care contacts. Participants were asked to report resource use at the baseline, 32, 64 and 96 weeks assessments. Resource use data are collected using and economic patient questionnaire (EPQ) which includes questions from the Client Service Receipt Inventory (CSRI) (26) interview and service use questionnaires used in previous mental health trials (27). At each assessment, participants were asked to recall their use of health and social care services since the previous scheduled assessment (or 6-months before baseline). The direct costs were estimated from this resource use data combined with the most recent published national unit costs available at the time of data analysis. These were the Department of Health Reference costs (28), the Unit costs of Health and Social Care produced by the Personal Social Services Research Unit, University of Kent (25), and the British National Formulary for the price year 2012-13. Each item of healthcare service use was assigned a cost by multiplying the quantity of service used with the average unit cost for that item. The ranges of unit costs of the different types of health and social care services used to calculate the costs of health and social care used by trial participants are reported in supplementary material (Table S1).

2.5 Missing data Multiple imputation (MI) was used to derive values based on all available data for a particular participant for the primary analysis of both costs and QALYs. MI of both costs and QALYs is increasingly recognised as an appropriate approach to deal with missing observation and missing follow-up data(29). All missing cost and utility data were treated as missing at random. Missing values were imputed for each time point, rather than as total values covering the whole follow up period. Costs were imputed by category and utility by individual EQ-5D domain, rather than as totals. This was so that all available data were used to inform the imputed values. For example, participants may have only been missing data relating to a single category of healthcare use, and so imputation was used to fill in the blanks. All available cost and outcome data for a particular participant was used to impute missing data. The imputations were conducted in STATA version 13, using predictive mean matching and sequential chained equations. The variables included in the models were selected on the basis of potential predictive ability identified from descriptive and regression analyses of the pooled baseline and follow up data (29).

The following variables were included in the imputation models: age; gender; whether or not the participant had undertaken further or higher education; whether the participant was currently in employment or not; whether the participant was white British or not; whether the participant was married or not; whether the participant lived alone or not; number of previous bipolar episodes; type of bipolar diagnosis; treatment allocation (PEd vs. PS); treatment group and wave; number of group sessions attended; last assessment attended; relapse (yes/no); time to relapse.

Participants with missing data at all time points were excluded from the imputation and subsequent analyses. Missing *total* cost values (incurred during the 96-week follow-up period) were estimated passively from the imputed costs at each follow-up assessment, missing utility values were estimated passively from imputed values for each EQ-5D item.

2.6 Approach The economic evaluation considered costs and outcomes incurred/received by health and social care agencies and patients; these are the key components of the NICE preferred Health and Personal Social Services (PSS) perspective (25). The setting was primary care in England. The time horizon was 96 weeks which reflects the need for longer term follow-up to identify relapse and recovery. As the time frame for the study was less than 2 full years, future costs and outcomes were not discounted (30).

To help decision-makers compare competing interventions, the results of cost-effectiveness analyses are reported as a ratio of how cost-effective one intervention is compared to another. This ratio summarises both how much more (or less) health benefit an intervention provides and how much more (or less) it costs and indicates the cost to gain an additional unit of benefit beyond that gained from the comparator. The primary measure for this economic analysis was the incremental cost-effectiveness ratio (ICER), estimated as:

$$\frac{Cost_{PEd} - Cost_{PS}}{QALY_{PEd} - QALY_{PS}}$$

As the primary measure of the economic analysis (ICER) is a ratio, rather than a point estimate cost-effectiveness acceptability analysis was used to assess the level of uncertainty in the data. Net costs

($Cost_{PEd}$ - $Cost_{PS}$) and net QALYs ($QALY_{PEd}$ - $QALY_{PS}$) were each derived using an ordinary least squares (OLS) regression model with total cost/utility as the dependent variable and treatment allocation as the independent variable. The primary analysis was adjusted for baseline characteristics of the sample.

2.6.1 Within-trial analysis For the PEd versus PS data evaluation the estimates of incremental costs and outcomes from the regression were bootstrapped to simulate 10,000 pairs of net cost and net outcomes. These simulations were used to estimate the *probability* that the PEd intervention is cost-effective compared with PS and to generate cost-effectiveness acceptability curves (CEACs). CEACs graphically depict the level of uncertainty around the results and show the probability that PEd is the more cost-effective option compared to PS at different levels of willingness to pay for each QALY gained (within the 10,000 bootstrapped net cost and QALY pairs). This takes a Bayesian approach to estimating the likelihood that the intervention is cost-effective and avoids hypothesis-testing and risk of a Type II error. In line with this approach no statistical tests of differences in mean costs or outcomes were conducted, although 95% confidence intervals around the differences are presented.

*2.6.2 Model-based analysis*_For the economic model comparing PEd with TAU, Monte Carlo simulation (10,000 iterations) using TreeAge software was used to estimate the net costs, QALYs and ICER. Probabilistic sensitivity analysis was used to estimate the likelihood that PEd was cost-effective compared with TAU and generate cost-effectiveness acceptability curves. This required that all parameters were entered into the model as distributions. Beta distributions for integers were used for the probability data. Gamma distributions were used for the cost, utility and time data. The cost and time distributions were constrained to a minimum value of zero.

2.7 Within trial analysis The within-trial analysis used an intent to treat approach and included all participants randomised to start therapy in both trial groups. Net costs and QALYs used to calculate the ICER for the primary economic analysis were derived using OLS regression to adjust for the baseline covariates identified as potential predictors of future costs and outcomes:: gender and age; study wave; number of previous episodes of BD; type of bipolar diagnosis (1 or 2); HADS-anxiety and HADS-depression scores; ethnicity (white British or not); employment status (working or not); level of education (further/higher education or not); living situation (living alone or not). These variables were used as covariates for all the primary and secondary analyses of costs, QALYs and cost-effectiveness. In addition, the baseline SF-6D utility score and EQ-5D visual analogue scale (VAS) score were used as covariates in the analyses of QALY data and baseline costs were used in the analyses of cost data.

Sensitivity analyses were used to assess the impact of uncertainty and the robustness of the results to changes in the measure of health benefit (time to bipolar episode) and missing data approach (MI versus complete case analysis). Sub-group analyses explored (rather than tested) whether the cost-effectiveness of PEd might vary according to the characteristics identified from clinical and qualitative analyses (as well as descriptive economic analyses) as potentially having an impact on costs or benefits associated with PEd or PS: the type of BD (I or II), number of (PEd/PS) sessions attended, and whether the participant completed the 96-week SCID assessment.

2.8 Decision tree and probabilistic simulation model A simple economic model that combined an

initial decision tree and a Markov model was constructed (Figure 1) to compare PEd to TAU (as opposed to PS). The initial decision tree structure was based on the care pathways used in the trial design and incorporates the distribution of participants by number of group sessions attended. Decision trees are simple and transparent, clarifying the options of interest. The Markov section of the model represents likelihood and time to first relapse following treatment, to reflect the primary objective of the PEd intervention. Markov models handle both costs and outcomes intuitively which makes them a powerful tool in economic evaluation (31). They are particularly useful for modelling chronic conditions with fluctuating severity, such as depression, over time. The time horizon was 96 weeks, split into three cycles of 32 weeks.

The target population for the economic model was people with bipolar 1 or 2 disorder at increased risk of further relapse (an episode in the last 24 months). Data from the trial sample of participants were used to represent this population.

2.8.1 Parameter estimation Parameters used in the model are summarised in Table 2. The following data were observed during the RCT for participants randomised to receive PEd: probability of attending group sessions; probability of first relapse in the follow-up period; costs of PEd; cost per week of relapse/no relapse, estimated as the average cost per assessment divided by 32 weeks (time between assessments); the average utility associated with relapse/no relapse; time to relapse and time following relapse.

It was assumed that the cost per week of relapse/no relapse, the utility associated with the relapse and no relapse states would be the same for both PEd and TAU. The probability of relapse and time to relapse for PEd (10-25%) were varied systematically around the values observed for PEd in the clinical analysis (14) to determine how effective PEd would have to be at delaying/preventing relapse, compared to TAU, to be cost effective (threshold analysis). The utility and cost parameters for the relapse and no relapse states were estimated from the multiple imputation data generated for the economic evaluation.

3. Results

The demographic characteristics of the sample are reported in full alongside the clinical effectiveness results (14). In summary, the mean age of the sample was 45 years, 58% were female, 91% were of white ethnicity, 80% had Type 1 BD, and over 50% had had 20+ previous bipolar episodes. 153 participants were randomised to receive PEd and 151 participants to receive PS.

A breakdown of the intervention costs by wave and study centre is presented in supplementary material (Table S2). Training, session delivery, and supervision costs were equivalent for PEd and PS, (total for each of PEd and PS - training: £14,896, sessions: £51,282, supervision: £2,080). PEd was more expensive than PS because of the in-depth manuals distributed to PEd participants (total cost £616). The total cost to deliver 21 sessions (one wave) of PS was £6,206 and £6,261 for PEd (calculated by dividing the total intervention cost by the number of waves (i.e. 11)). . The intervention design was such that the cost to deliver each wave of sessions were incurred whether or not all participants attended all sessions. The estimated intervention cost per participant was calculated by dividing the cost per wave by the average number of planned participants per wave. The mean number of planned participants per wave was 14 and so the cost per participant was

estimated as the mean cost to deliver 21 sessions divided by the planned group size i.e. £6261/14=£447 for PEd and £6206/14=£443 for PS. Given the similar costs of the two interventions, additional costs incurred in the PEd participants indicates higher use of health and social care services. Mean values for the main categories of healthcare costs and utility for PEd and PS are reported in supplementary material (Table S3).

The pattern of available cost and utility data across the different assessments are summarised in Table S4. The proportion of the sample that it was possible to calculate costs or utilities for was 91% and 85% at baseline and decreased at each subsequent assessment. The proportion of the sample with at least partial data recorded for resource use or utilities was marginally higher at baseline, 98% and 88% respectively. At the final assessment (week 96) 57% had full data recorded for resource use and 33% had full EQ-5D data allowing estimation of a utility value. Additionally, 65% (PS) and 71% (PEd) participants had partial data about service use at the final assessment. Over half the sample (54%) had both resource use and utility data at 2 or more time points and a third (34%) had these data at 3 or 4 time points. Around a quarter (23%) of the sample had resource use and utility data at both baseline and final assessment. Overall, cost data were more complete than utility data with similar proportions of available data for PEd and PS participants.

3.1 PEd versus PS (within-trial analysis) Table 1 reports net costs and QALYS, estimated from regression analysis of imputed data (after controlling for the effect of baseline characteristics, trial implementation, and treating costs and QALYs as independent of each other). There was a small QALY gain of 0.023 (95% CI, 0.001-0.056 associated with the PEd intervention and an additional cost (healthcare resources used) of £1098 (95% CI, £252-£1943). The point estimate of the net cost per QALY gained is relatively high at £47,739 (£1098/0.023 QALY), which is above the commonly reported NICE threshold of £20,000 to £30,000 per QALY gained (18). There is some uncertainty associated with this cost per QALY estimate. This uncertainty is illustrated in Figure 2 (panel a), which presents a scatterplot of the 10,000 bootstrapped pairs of net cost and QALY data (cost-effectiveness plane). The cost-effectiveness plane shows that the net costs are mostly scattered above the horizontal axis, suggesting that overall PEd is likely to cost more than PS. The net QALY points on the scatterplot are mainly to the right of the vertical axis, suggesting that PEd is more likely than PS to produce a health benefit. The uncertainty is also reflected in the probability that PEd is more cost-effective than PS (Table 1). The cost-effectiveness acceptability curve (CEAC) presented in Figure 2 (panel b) illustrates that compared with PS, the PEd intervention is likely to be cost-effective in 50% or more cases if decision-makers are willing to pay £37,500 or more to gain one QALY. This value is a measure of the uncertainty around the ICER, commonly reported in cost-effectiveness evaluations, and is akin to reporting a 95% CI alongside a point estimate.

3.1.1 PEd versus PS: Sensitivity and sub-group analyses PEd is more likely to be cost-effective than PS in the following scenarios (Table 1): complete case analysis (no multiple imputation of missing data); over 64 weeks (rather than 96 weeks); alternative measures of health benefit (relapse avoided; relapse free years). Sub-group analysis exploring the impact of Type of bipolar diagnosis indicates that PEd was more likely to be cost effective compared to PS in participants with a Type 2 bipolar diagnosis. PS dominates PEd (lower costs and higher QALYs for PS) in the following scenarios: fewer than 8 previous bipolar episodes; fewer than 16 group sessions attended; when participants did not complete SCID at 96-week assessment. It is important to note that these analyses are

exploratory and the sample size available for each analysis may not be sufficient to identify important differences. The cost-effectiveness acceptability curves for selected sensitivity and sub-group analyses are summarised in Figure 2 (panels c and d). The full results for all sensitivity and sub-group analyses conducted are reported in supplementary material (Tables S5-6).

3.2 PEd versus TAU (model-based analysis) Table 2 summarises the decision tree events and structure and describes the probability data inputs to the economic model and the cost, utility and time data used, derived from the within-trial analyses.

The net costs and QALYs of PEd compared with TAU are shown in Table 3, while Figure 3 gives the cost-effectiveness acceptability curves for the analyses. The analyses demonstrate that if PEd reduces the probability that a person has a relapse by 15% (regardless of time to relapse) compared to TAU, then the net cost to gain one QALY from PEd (£9700) falls well below the £30,000 threshold. If the probability of relapse is 25% higher in TAU than PEd, then PEd dominates TAU (lower costs, higher QALYs for PEd). A systematic review (12) indicated that for TAU, the probability of relapse over an average of 60 weeks is 0.70. At 60 weeks, the probability of relapse observed for the PEd participants in our study was 0.50 i.e. probability of relapse is 20% higher for TAU.

The clinical analysis showed that the median time from baseline to next bipolar episode was 67 weeks (95% CI 37–91) in the psychoeducation group versus 48 weeks (95% CI 31–66) in the peer-support group (14). Results from the economic model suggest that if PEd reduces the probability that a person has a relapse by 10% *and* increases the time to relapse by 10% compared to TAU, then the net cost to gain one QALY (£10,300) is again lower than £30,000. If the probability of relapse is 15% higher and time to relapse is 15% lower in TAU than PEd, then PEd dominates TAU (lower costs, higher QALYs for PEd). The average time to relapse observed in our study was around 30 weeks (95% CI, 24-35) for PEd. Previous studies suggest that the average time to relapse under TAU is around 20 weeks (12) (i.e. 33% sooner).

Even though the ICERs reported in Table 3 are below £20,000, the estimated probability that PEd is cost-effective compared to TAU if decision-makers are willing to pay £20,000 or even £30,000 are not much higher than chance (0.50). This reflects the uncertainty around the ICERs.

4. DISCUSSION

Primary analysis suggested a net additional cost for the PEd intervention and a small QALY gain compared to the active PS control intervention. There is a 35% probability that PEd is cost-effective, compared to PS if decision makers are willing to pay £30,000 to gain one QALY. The PEd intervention, compared with PS, was only likely to be cost-effective if decision-makers are willing to pay at least £37,500 to gain one QALY. The incremental cost to gain one relapse-free year was £8382 and to avoid one relapse was £10,765. Decision-makers must be prepared to pay at least these amounts for the respective outcomes for PEd to be potentially more cost-effective than PS. However PS of the intensity used for the control arm in this trial is not available in routine care and so comparison with TAU may be more meaningful for decision-makers.

The results of the economic model suggest that if PEd is more effective at reducing the probability of relapse or delaying relapse compared with TAU then PEd could potentially be cost-effective. The

difference in effectiveness of PEd and TAU required for PEd to be cost-effective is smaller than the difference in relapse times/rates observed in our study (PEd) and those reported in a systematic review (TAU) (12). However the economic model analyses indicated that the 95% CIs for the net costs and QALYs were wide and included zero, indicating a high level of uncertainty around these estimates.

The higher costs observed for participants in the PEd group arose from healthcare services used, rather than the intervention (which cost £4 per participant more than PS). One element of PEd encouraged appropriate use of healthcare services which may explain this observation. However healthcare utilisation did not show any clear pattern over the 96 week follow-up to support or contradict this. Evaluation of healthcare utilisation over a longer time period following PEd may offer further explanation.

4.1 Strengths and limitations The within-trial economic analysis was subject to the same strengths and limitations as described in detail for the clinical evaluation of PEd (14). In brief, there was high retention during follow-up, good internal and external validity, inconsistent attendance at treatment groups, and possible floor effects as an artefact of recruiting only participants with low levels of depression and mania. An additional issue that is relevant for the economic analysis is the low level of participants with complete cost and utility data.

Although there was a high follow up rate for the primary outcome of the trial, the number of participants completing the secondary outcome measures (including the EQ-5D) and service use questionnaire was much lower, in part reflecting a focus on ensuring that the primary outcome measure was completed. As demonstrated in Table S4, the extent of missing data differed between service use and estimated costs (57% of participants had complete cost data at the week 96 assessment) and QALYs (33% had complete 96-week data). This was due in part to partially completed EQ-5D and healthcare utilisation surveys. This adds uncertainty to the results of the primary and sensitivity analyses of the costs and outcomes. Multiple imputation was used to impute missing values to "fill in the blanks" while still using all available data. The regression models used to impute missing data were based on initial descriptive analyses and regression analyses to identify key baseline and follow-up variables that were associated with either costs or outcomes. These were included in the chained equations to iteratively impute missing cost categories and utilities. This goes some way to controlling for the influence of observed variables on the missing data. However, the influence of unobserved variables could not be taken into account, and it is unclear the extent to which this was important and affects the assumption that missing data were missing at random.

Given the level of missing data, this number of MI datasets for the cost estimates is likely to be 98% efficient with a 2% loss of power. For QALY estimates, 10 MI datasets is likely to be 94% efficient with a 9% loss of power. Overall the efficiency of 10 MI datasets for the ICER estimates is likely to be lower (89%-93%) and the loss of power higher (14%) (32). Using the rule of thumb suggested by White et al (33), we would need 43 MI datasets for costs, 67 MI datasets for QALYs and 83 MI datasets to ensure that the results from the MI analyses are stable. However, the available data approach used would increase the computational burden substantially.

The economic analyses used OLS regression to take account key baseline characteristics that may affect costs and QALYs. In cases where clustering of data occurs (e.g. by treatment group) multi-level analyses may be a better approach for reliable and unbiased estimation of the cost and QALY coefficients and standard errors. The primary clinical analysis found no evidence of clustering by therapy group. The exploratory sub-group analyses identified possible effects that could be assessed in future using multi-level modelling approaches.

QALYs were used as the measure of health benefit for the primary analysis and most of the sensitivity and sub-group analyses. This enables comparison between different disorders which is relevant for policy makers and commissioners, who have to consider the distribution of limited budgets between different health care services. However, there is a possibility that a generic health status measure may not be sufficiently sensitive to identify important clinical changes in participant's mental health or bipolar status and so could underestimate the benefits of PEd. However baseline utility values derived from the EQ-5D in this sample correlated well with the clinical measures used in the trial.

The costs included in the economic evaluation were limited to the direct costs of providing health and social care, in accordance with the perspective taken. This will underestimate the total costs of bipolar disorder for the trial participants. If there are differences in the effectiveness of the PEd or PS interventions that lead to differences in employment between the two groups, then this will also affect the net costs associated with PEd. If employment is higher in the PEd group, then the net cost of the PEd group will be slower than estimated, and the likelihood that PEd is cost–effective may increase, and vice versa.

An assessment of the likely cost-effectiveness of PEd compared with TAU is important for commissioners and service providers. The PS intervention used as the control treatment in our study involved a closed group of only people with BD meeting with health professionals and a peer facilitator with an agenda for each session previously set by consensus by the group. This type of group support is not available in UK routine care. Typically PS is delivered by peers in a open format with people with BD and their carers or families attending as frequently as they like and usually with no set agenda for each session (http://www.bipolaruk.org/Pages/Category/support-groups). However there was limited evidence from the published literature to inform the structure of a full economic model that compares PEd with TAU or estimate the costs and QALYs associated with events in the TAU arm of the model. The advantage of populating the modelling with data from the RCT in relation to PEd is that it introduces less heterogeneity and uncertainty than had data for both PEd *and* TAU been synthesised from several sources. However this also means that the scope of the model and relevance of the model results are constrained by the trial design and the people with BD recruited into the main RCT.

If as seems likely, commissioners and service providers wish to focus on relapse-free years, then PEd is much more likely to be cost-effective than PS, with less uncertainty. Whether the NICE thresholds for cost per relapse-free year gained would be as high as those for cost per QALY gained is unclear

4.2 Implications Sensitivity and sub-group analyses were conducted to explore whether aspects of the trial and economic evaluation design may affect the likelihood that PEd is cost-effective. This approach can help to identify issues for policy and indicated a number of areas for further research. These analyses also explored whether there were specific groups of participants who may benefit more from PEd. However, it is important to note that these analyses are exploratory, and the sample size available for each analysis may not be sufficient to identify important differences. Additionally, the number of these exploratory analyses increases the chance of finding a difference between the exploratory analysis and the primary analysis because of the number of analyses, rather than because an important difference exists.

The results were sensitive to the length of follow-up. PEd was more likely to be cost-effective at the 64-week assessment than the 96-week assessment- driven by a difference in net costs (rather than QALYs). Further work to identify the trajectory of costs and QALYs during the intervention phase and over the longer term is needed. This will help to understand whether the change found in our study reflects what would happen is routine practice.

Sub-group analysis of the economic data indicated that PEd might be cost-effective compared with PS for people with Type 2 BD and those who attended 16 or more sessions. The sub-group analyses also indicate that the number of sessions attended affected the *relative* cost-effectiveness of PEd. This highlights the need for careful identification of people who are likely to engage with and benefit from the intervention. Further work to understand the influence of a participants' overall health on decisions to participate in PEd or PS is needed. This would help to target the intervention at participants able and willing to engage in the group intervention and/or improve access to group interventions.

4.3 Conclusion In conclusion, PEd is associated with a modest improvement in health status and higher costs than PS. There is a high level of uncertainty in the data and results.

Acknowledgements: We thank Bipolar UK, Mind, and Mood Swings for helping to publicise the study; Manchester Health and Social Care NHS Trust for hosting the PARADES programme; the University of Nottingham for providing sponsorship and clinical trials unit support; and the Spectrum Centre, University of Lancaster, and University of Manchester for additional support. In particular we acknowledge the contributions of the NIHR PARADES Psychoeducation Study Group, without whom the study would not have been possible: Rebecca Anderson, Marcus Barker, Lucy Bateman, Alison Beck, Nancy Black, Lisa Brown, Phillip Byrne, Jane Fisher, Lynda Fretwell, Lorraine Getten (now Ward), Kay Hampshire (now Glint), Paul Hammersley, Claire Hilton, Jelena Jovanoska, Steven Kendall, Dawn Knowles, Rachel Lambourne, Brian Langshaw, Susan Lomas, Natasha Lyon, Deborah Mayes, Elly McGrath, Sian Newman, Kirsten Nokling, Puru Pathy, Kathryn Reeveley, Kirsty Stevenson, Katherine Taylor, Karthik Thangavelu, Elizabeth Tyler, and Emma Weymouth. We also thank the participants in the study.

Contributors: RM, FL, LD, and SJ obtained funding for the study. RM led the study in the East Midlands and FL led the study in the North West. LR was the trial coordinator. RM, FL, and SJ supervised the interventions, and supervised data collection with LR. EC and DN wrote the economic analysis plan, EC cleaned and analysed the data under supervision from LD. EC led the writing of the manuscript and revised subsequent drafts. All authors contributed to and approved the final version of the manuscript.

Table 1 Net costs and QALYs of group psychoeducation (PEd) compared to group peer support (PS), multiple imputation data, adjusted for baseline characteristics* and bootstrapped

	Net cost (95% CI)	**Net QALY (95% CI)**	**Net cost per QALY gained**	**Probability PEd cost effective if willing to pay £30K to gain 1 QALY**
Primary analysis	£1098 (252; 1943)	0.023 (0.001; 0.056)	£47,739	0.35
		Sensitivity analyses		
Type of imputation				
Complete case analysis (n=55)	-£588 (-4734; 3558)	0.049 (-0.177; 0.274)	PEd dominates	0.83
Shorter length of follow up				
Baseline to 64 weeks (n=280)	£698 (209; 1187)	0.027 (0.006; 0.048)	£25,852	0.63
Alternative measures of health benefit				
	Net cost (95% CI)	**Net time relapse free (years) (95% CI)**	**Net cost per relapse free year gained**	**Probability PEd cost effective if willing to pay £30K to gain 1 relapse free year**
Relapse free years (n=287)	£1098 (252; 1943)	0.131 (0.08; 0.182)	£8,382	0.99
	Net cost (95% CI)	**Net number of relapses avoided (95% CI)**	**Net cost per relapse avoided**	**Probability PEd cost effective if willing to pay £30K to avoid 1 relapse**
Relapse avoided during follow-up (n=287)	£1098 (252; 1943)	0.102 (0.67; 137)	£10,765	0.99
		Sub-group analysis		
	Net cost (95% CI)	**Net QALY (95% CI)**	**Net cost per QALY gained**	**Probability PEd cost effective if willing to pay £30K to gain 1 QALY**
Participants with Type 2 bipolar diagnosis (n=58)	£375 (-107; 857)	0.129 (0.077; 0.181)	£2907	1.00
0-15 PEd or PS sessions attended (n=169)	£955 (332; 1577)	-0.003 (-0.040; 0.043)	PS dominates	0.07
16+ PEd or PS sessions attended (n=118)	-£844 (-1997; 308)	0.080 (0.036; 0.124)	PEd dominates	0.99

*gender, age, study wave, number of previous episodes of BD, type of bipolar diagnosis (1 or 2), HADS-anxiety and HADS-depression scores, ethnicity, employment status, level of education, living situation, SF-6D utility score, EQ-5D thermometer, pre-baseline healthcare costs

Table 2 Economic model parameters: probability of events, costs, utility and time to relapse

Model branch	Event sequence		Probability (95 CI%)
1	Attends at least one session		0.78 (0.71-0.74)
1.1	Attends 2-9 sessions		0.19 (0.13-0.27)
1.1.1	First relapse, baseline-week 32		0.61 (0.39-0.79)
1.1.2	No first relapse, baseline-week 32		0.39 (0.21-0.61)
1.1.2.1	First relapse, week 33-64		0.11 (0.01-0.55)
1.1.2.2	No first relapse, week 33-64		0.89 (0.45-0.99)
1.1.2.2.1	First relapse, week 65-96		0.13 (0.01-0.60)
1.1.2.2.2	No first relapse, week 65-96		0.87 (0.40-0.99)
1.2	Attends more than 9 sessions		0.81 (0.73-0.87)
1.2.1	Attends 10-15 sessions		0.31 (0.22-0.41)
1.2.1.1	First relapse, baseline-week 32		0.43 (0.27-0.72)
1.2.1.2	No first relapse, baseline-week 32		0.57 (0.38-0.73)
1.2.1.2.1	First relapse, week 33-64		0.18 (0.05-0.45)
1.2.1.2.2	No first relapse, week 33-64		0.82 (0.55-0.95)
1.2.1.2.2.1	First relapse, week 65-96		0.21 (0.06-0.52)
1.2.1.2.2.2	No first relapse, week 65-96		0.79 (0.48-0.94)
1.2.2	Attends more than 15 sessions		0.69 (0.59-0.78)
1.2.2.1	First relapse, baseline-week 32		0.33 (0.23-0.35)
1.2.2.2	No first relapse, baseline-week 32		0.67 (0.55-0.77)
1.2.2.2.1	First relapse, week 33-64		0.13 (0.06-0.27)
1.2.2.2.2	No first relapse, week 33-64		0.87 (0.73-0.94)
1.2.2.2.2.1	First relapse, week 65-96		0.23 (0.12-0.39)
1.2.2.2.2.2	No first relapse, week 65-96		0.77 (0.61-0.88)
2	Attends 0 sessions, continue as control		0.22 (0.16-0.20)
Parameter	**PEd**	**PS**	**All**
	Mean (95% CI)	**Mean (95% CI)**	**Mean (95% CI)**
Utility			
No relapse	0.703 (0.635-0.771)	0.669 (0.595-0.743)	0.689 (0.638-0.739)
Relapse	0.592 (0.530-0.654)	0.588 (0.524-0.653)	0.590 (0.543-0.637)
Cost, £			
Intervention	450 (446-454)	452 (448-456)	451 (448-454)
No relapse	1554 (1109-1999)	1331 (850-1811)	1458 (1117-1799)
Relapse	3154 (2103-4205)	2736 (2035-3436)	2933 (2318-3549)
Weeks to relapse (within cycle)			
Week 0-32	14.47 (12.11-16.82)	13.04 (11.02-15.05)	13.74 (12.20-15.28)
Week 33-64	18.03 (12.46-23.60)	17.05 (13.69-20.41)	17.39 (14.52-20.27)
Week 65-96	18.85 (14.14-23.56)	18.86 (13.68-24.03)	18.85 (15.44-22.26)
Week 0-96	29.63 (24.21-35.06)	28.54 (23.76-33.32)	29.06 (25.49-32.63)

Table 3 Economic model: Net costs and QALYs of PEd compared with TAU

	Net cost, (95% CI)	Net QALY, (95% CI)	Net cost per QALY, £	Probability PEd cost effective		
				WTPT = £20K per QALY	WTPT = £25K per QALY	WTPT = £30K per QALY
Probability of relapse higher in TAU than PEd[1]						
15% higher	97 (-6613; 4890)	0.010 (-0.266; 0.269)	9700	0.48	0.50	0.51
20% higher	18 (-7404; 5065)	0.014 (-0.277; 0.290)	1286	0.50	0.51	0.52
25% higher	-60 (-8142; 5136)	0.017 (-0.292; 0.312)	PEd dominates	0.51	0.52	0.53
Probability *and* time to relapse[2]						
Probability 10% higher, time 10% lower	£103 (-5912; 4320)	0.010 (-0.233; 0.240)	£10,300	0.48	0.49	0.50
Probability 15% higher, time 15% lower	-£18 (-6710; 4123)	0.015 (-0.233; 0.263)	PEd dominates	0.50	0.51	0.52
Probability 20% higher, time 20% lower	£-141 (-7558; 4135)	0.020 (-0.239; 0.288)	PEd dominates	0.52	0.53	0.54

[1]probability of relapse in PEd = 0.58 (95% CI 0.50-0.66)

[2]mean time (weeks) to relapse in PEd = 29.63 (95% CI 24.21-35.06)

WTPT = willingness to pay threshold

Figure 1 Economic model: decision tree and Markov model structure

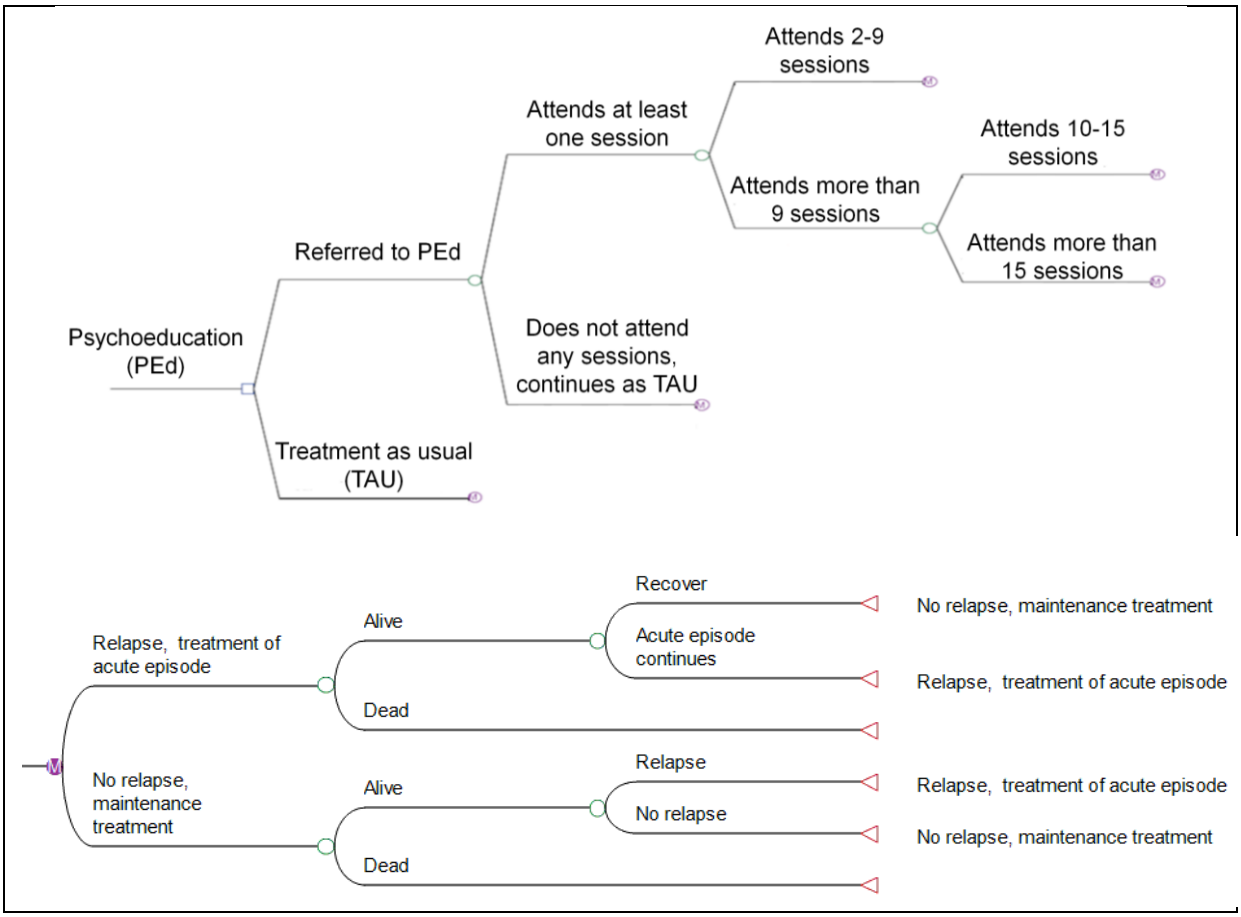
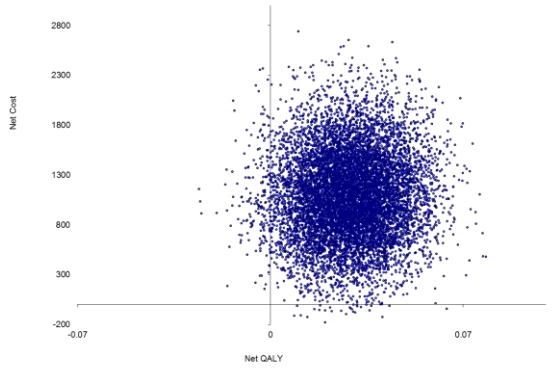
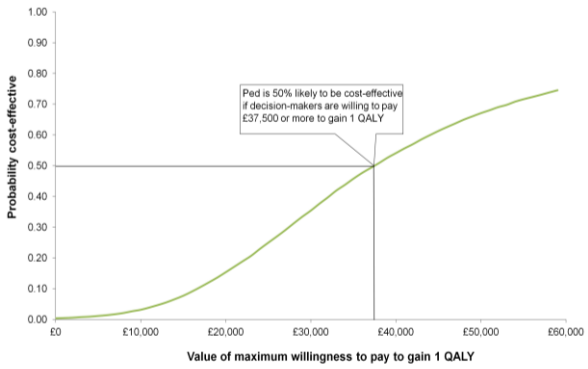


Figure 2 (a) Cost effectiveness plane of bootstrapped net cost and QALY pairs (primary analysis); (b) cost effectiveness acceptability curve (primary analysis); (c) cost effectiveness acceptability curve (complete cases, clinical outcome measures); (d) cost effectiveness acceptability curve (length of follow-up, number of sessions attended, bipolar type)

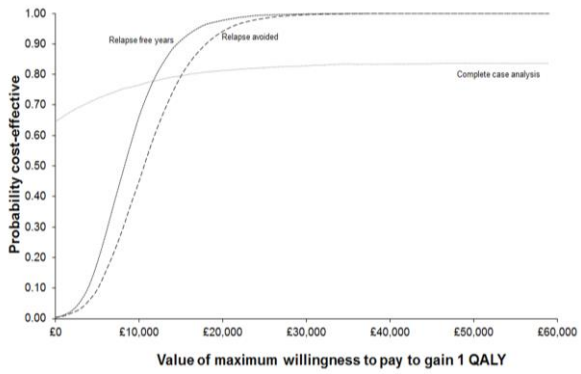
(a)



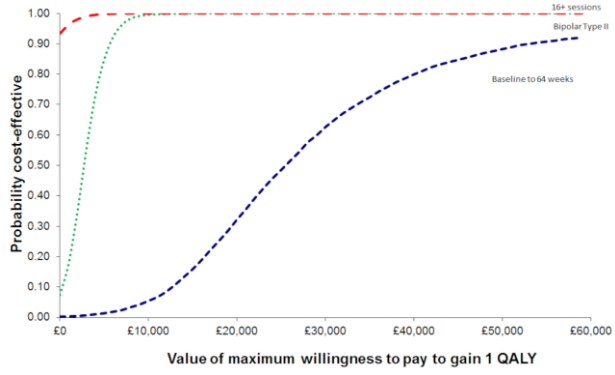
(b)



(c)



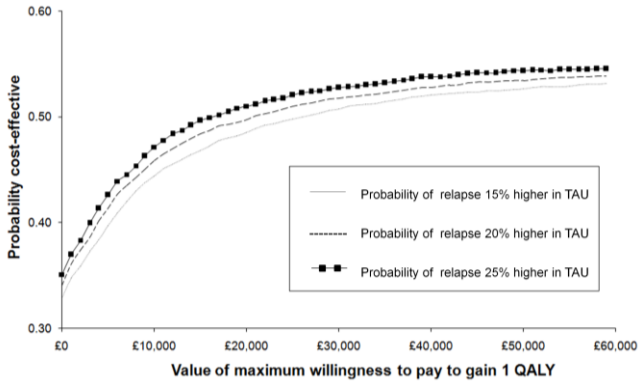
(d)



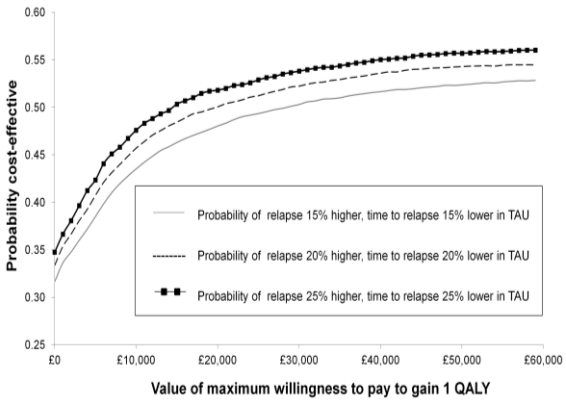
Data represented in figures are from multiple imputed datasets and have been adjusted for the following covariates: gender and age; study wave; number of previous episodes of BD; type of bipolar diagnosis (1 or 2); HADS-anxiety and HADS-depression scores; ethnicity (white British or not); employment status (working or not); level of education (further/higher education or not); living situation (living alone or not). Baseline SF-6D utility score and EQ-5D thermometer were used as covariates in the analyses of QALY data only and baseline costs were used in the analyses of cost data only.

Figure 3 Cost effectiveness acceptability curves, willingness to pay to gain 1 QALY, comparison with TAU: (a) probability of relapse; (b) time to and probability of relapse

(a)



(b)



Supplementary material

Table S1 Mean (range) unit costs of health and social care services

Service type	**Range of unit costs (£)**	**Source**
Psychiatric hospital inpatient admission	348 per day (acute psychiatric ward) 334 per day (rehabilitation ward) 736 per day (psychiatric intensive care ward)	NHS Reference costs (28)
Other hospital inpatient admission	491 per day (general medical ward) 1,650 per day (Ear nose and throat, day case admission)	NHS Reference costs (28)
Psychiatric hospital outpatient care	221 per visit (adult mental illness) 102-216 per visit (other psychiatric outpatient visits) 145-191 per visit (psychotherapy, psychology) 184 per visit (crisis team) 24.90 per visit (depot clinic)	NHS Reference costs (28)
Other hospital outpatient admissions	108 per visit (type of visit not specified, costed as average of all outpatient visits) 26-215 per visit (other physical health care)	NHS Reference costs (28)
Tests and diagnostics	3-4 per test (haematology, phlebotomy) 28-177 per test (x-rays, scans, MRI, biopsy)	NHS Reference costs (28)
General practice		
GP	37.50 (per surgery visit) 94.50 (per home visit)	PSSRU Unit costs of health and social care (25)
Practice nurse	12.40 per visit Single estimate, no range	PSSRU Unit costs of health and social care (25)
District nurse, health visitor, midwifery	59.58 per visit Single estimate, no range	PSSRU Unit costs of health and social care (25)
Other primary physical health care	25-115 per visit	NHS Reference costs (28) and PSSRU Unit costs of health and social care (25)
Community mental health/social care	13.50 per visit (Home help/Care worker/health care assistant/clinical or care support worker/link mental health worker/support time and recovery worker) 68-135 per visit (Community Psychiatric Nurse/Case Manager, care coordinator/CMHT) 58-123 per visit (counselling and therapy) 204.50 per visit (psychiatrist) 91-204 (other miscellaneous services) 207 per visit (social worker) 13.50-104 per visit (other miscellaneous services)	NHS Reference costs (28) and PSSRU Unit costs of health and social care (25)

Table S2 Cost[a] of the psychoeducation (PEd) and peer support (PS) interventions

Wave, location, year started	PEd			PS		
	Planned no. participants	No. (%) attend 1+ sessions	Cost/planned participant, £	Planned no. participants	No. (%) attend 1+ sessions	Cost/planned participant, £
Wave 1 Manchester 2009	12	10 (83)	522	12	10 (83)	517
Wave 2 Manchester 2010	17	13 (76)	368	15	12 (80)	414
Wave 3 Barrow 2010	10	9 (90)	626	11	11 (100)	564
Wave 4 Carlisle 2010	17	15 (88)	368	17	16 (94)	365
Wave 5 Preston 2010	16	12 (75)	391	16	14 (88)	388
Wave 6 Preston 2011	14	14 (100)	447	15	11 (73)	414
Wave 7 Nottingham 2009	14	13 (93)	447	11	11 (100)	564
Wave 8 Mansfield 2010	11	11 (100)	569	10	7 (70)	621
Wave 9 Boston 2010	11	8 (73)	569	13	10 (77)	477
Wave 10 Chesterfield 2011	14	11 (79)	447	14	12 (86)	443
Wave 11 Nottingham 2012	17	17 (100)	368	17	15 (88)	365

[a]Costs reported in table are rounded to nearest £

The cost for each wave of 21 sessions was calculated as: total intervention cost (training, delivery, supervision) for treatment type/number of waves (i.e. 11)

The cost per planned participant was calculated as: wave cost/number of planned participants

Mean intervention cost per participant was calculated as: cost per planned participant*mean number of planned participants (i.e. 14)

Unit costs for facilitators/trainers (derived from PSSRU Unit Costs of Health and Social Care were: Consultant Psychiatrist, £101/hour; Clinical Psychologist, £59/hour; Trainee Psychiatrist, £56/hour; Therapist, £42/hour; Service user, £13/hour.

Table S3 Mean utility values and costs of healthcare services used by treatment group (imputed data)

	PEd (*n* = 153)	PS (*n* = 151)
EQ-5D Utility values	**Mean (95%CI)**	**Mean (95%CI)**
Baseline	0.69 (0.64- 0.74) n=133	0.68 (0.63-0.73) n=125
32-week assessment	0.70 (0.64-0.77) n=75	0.63 (0.57-0.70) n=86
64-week assessment	0.66 (0.59-0.74) n=66	0.62 (0.53-0.71) n=60
96-week assessment	0.63 (0.53-0.72) n=49	0.70 (0.64-0.77) n=50
Services used	**Mean cost (95%CI), £**	**Mean cost (95%CI), £**
6 months prior to baseline		
Hospital inpatient admission	1861 (842-2880) n=149	2233 (935-3531) n=144
Hospital outpatient care	609 (521-697) n=150	613 (486-739) n=144
General practice[a]	180 (138-222) n=152	168 (129-208) n=145
Other primary physical health care	58 (37-79) n=152	48 (27-68) n=144
Community mental health/social care	379 (228-530) n=148	416 (254-577) n=143
Prescription medications	461 (321-600) n=148	280 (198-363) n=144
Baseline to week 32		
Hospital inpatient admission	981 (136-1526) n=121	671 (226-1116) n=112
Hospital outpatient care	481 (391-571) n=121	502 (399-606) n=112
General practice[a]	187 (115-259) n=121	199 (118-280) n=112
Other primary physical health care	39 (18-60) n=121	25 (11-38) n=112
Community mental health/social care	364 (220-509) n=121	249 (143-355) n=111
Prescription medications	538 (375-701) n=113	360 (268-452) n=105
Week 32 to week 64		
Hospital inpatient admission	911 (0-2033) n=108	1120 (132-2017) n=100
Hospital outpatient care	370 (293-447) n=108	431 (338-523) n=100
General practice[a]	159 (118-200) n=108	134 (97-170) n=100
Other primary physical health care	14 (1-26) n=107	33 (11-56) n=100
Community mental health/social care	363 (176-550) n=108	295 (162-428) n=100
Prescription medications	504 (331-677) n=97	361 (260-461) n=88
Week 64 to week 96		
Hospital inpatient admission	1360 (23-2697) n=107	1386 (439-2332) n=95
Hospital outpatient care	352 (268-436) n=107	387 (291-483) n=95
General practice[a]	95 (78-111) n=107	125 (92-159) n=95
Other primary physical health care	15 (5-25) n=107	33 (4-62) n=95
Community mental health/social care	470 (276-664) n=107	343 (91-594) n=95
Prescription medications	652 (426-878) n=92	431 (327-535) n=84

[b]GP, practice nurse, district nurse

Table S4 Summary of available cost and utility data at different assessments

	PEd n (%) [*% partial data*][a]		PS n (%) [*% partial data*][a]		Both treatment arms (%)	
	Cost	**Utility**	**Cost**	**Utility**	**Cost**	**Utility**
Study time point						
Baseline	142 (93) *[99]*	133 (87) *[88]*	135 (89) *[98]*	125 (83) *[88]*	91%	85%
Week 32	112 (73) *[80]*	75 (49) *[49]*	100 (66) *[77]*	86 (57) *[57]*	70%	53%
Week 64	96 (63) *[71]*	66 (43) *[44]*	85 (56) *[68]*	60 (40) *[40]*	60%	41%
Week 96	91 (59) *[71]*	49 (32) *[33]*	82 (54) *[65]*	50 (33) *[34]*	57%	33%

Number of assessments with complete cost and utility data	**PEd n (%)**	**PS n (%)**	**Both treatment arms (%)**
0	23 (15)	23 (15)	15%
1	41 (27)	52 (35)	31%
2	37 (24)	26 (17)	21%
2+	*89 (58)*	*76 (50)*	*54%*
3	25 (16)	26 (17)	17%
3+	*52 (34)*	*50 (33)*	*34%*
4	27 (18)	24 (16)	17%
baseline and wk 96	39 (25)	31 (21)	23%

[a]Response to at least one domain on EQ-5D and cost for at least one category of resource use

Table S5 Net costs and QALYs of PEd compared to PS, sensitivity analyses, bootstrapped data, adjusted for baseline characteristics

Sensitivity analysis	Net cost, £ (95% CI)	Net QALY, (95% CI)	Net cost per QALY gained, £	Probability PEd cost-effective If willing to pay £30K to gain one QALY*
Primary analysis	1098 (252 to 1943)	0.023 (0.001 to 0.056)	47 739	0.35
Type of imputation				
Complete case analysis (*n* = 55)	-588 (-4734 to 3558)	0.049 (-0.177 to 0.274)	PEd dominates, QALY gain, net saving	0.83
Missing QALY data imputed by linear interpolation (*n* = 287)	1122 (272-1971)	0.024 (0.004-0.045)	46 750	0.23
Alternative measures of health benefit				
Relapse free years (*n* = 287)	1098 (252-1943)	0.131 (0.08-0.182)	8382	Probability PEd cost-effective if willing to pay £30K to gain one relapse-free year: 0.99
Relapse avoided (*n* = 287)	1098 (252-1943)	0.102 (0.68-136)	10 765	Probability PEd cost-effective if willing to pay £30K to avoid one relapse: 0.99
QALY estimated from SF6D utility values (*n* = 287)	1098 (252-1943)	0.013 (0.000-0.027)	84 461	0.07
Alternative cost measures				
Costs medications, exclude costs 'other' medicines (*n* = 289)	1008 (162-1854)	0.028 (0.001-0.056)	25 200	0.41
Costs exclude costs of PEd/PS (*n* = 287)	1124 (275-1974)	0.028 (0.001-0.056)	40 142	0.34

Table S6 Net costs and QALYs of PEd compared to PS, sub-group analyses, bootstrapped data, adjusted for baseline characteristics

Sub-group analysis	Net cost, £ (95% CI)	Net QALY, (95% CI)	Net cost per QALY gained, £	Probability PEd cost-effective (£30K/QALY gain)
Primary analysis	1098 (252 to 1943)	0.023 (0.001 to 0.056)	47 739	0.35
No. PEd or PS sessions attended				
<10 (*n* = 112)	149 (-555 to 852)	0.019 (-0.025-0.064)	7842	0.72
10+ (*n* = 175)	1266 (24-2507)	0.023 (-0.012-0.057)	55 043	0.25
16+ (*n* = 118)	-844 (-1997 to 308)	0.080 (0.036-0.124)	PEd dominates	0.99
1-15 (*n* = 138)	1598 (896-2301)	0.001 (-0.040-0.043)	1 598 000	0.02
0-15 (*n* = 169)	955 (332-1577)	-0.003 (-0.040-0.034)	PS dominates	0.07
0 (*n* = 31)	-649 (-2114 to 814)	0.056 (-0.049-0.161)	PEd dominates	0.92
No. previous episodes				
≥8 (*n* = 249)	1238 (270-2206)	0.030 (0.001-0.060)	41 267	0.34
1-7 (*n* = 38)	1267 (-451 to 2985)	-0.056 (-0.132-0.019)	PS dominates	0.04
No. previous episodes, No. sessions attended				
8+ episodes, < 16 sessions (*n* = 147)	1026 (385-1667)	0.012 (-0.028-0.052)	85 500	0.21
8+ episodes, 16+ sessions (*n* = 103)	-1109 (2504-287)	0.043 (-0.003-0.089)	PEd dominates	0.99
1-7 episodes, 16+ sessions (*n* = 15)	-713 (-1898-472)	0.266 (-0.201-0.733)	PEd dominates	0.99
1-7 episodes, <16 sessions (*n* = 23)	6271 (3716-8826)	-0.133 (-0.221 to -0.044)	PS dominates	0.00
Participant completed SCID at 96-week follow-up				
SCID score (*n* = 212)	983 (-25 to 1990)	0.038 (0.009-0.067)	25 868	0.61
No SCID score (*n* = 75)	470 (-273 to 1213)	-0.017 (-0.003 to 0.049)	PS dominates	0.18
Type 1 or 2 bipolar diagnosis				
Type 1 (*n* = 229)	1183 (177-2189)	0.006 (-0.025 to 0.038)	197 167	0.009
Type 2 (*n* = 58)	375 (-107 to 857)	0.129 (0.077-0.181)	2907	1.00

Reference List

(1) Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2013;380(9859):2163-96.

(2) McCrone PR, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. Paying the price: the cost of mental health care in England to 2026. King's Fund; 2008.

(3) National Collaborating Centre for Mental Health, National Institute for Clinical Excellence. Bipolar disorder: The NICE guideline on the assessment and management of bipolar disorder in adults, children, and young people in primary and secondary care - CG185. National Institute for Clinical Excellence; 2014.

(4) Morselli PL, Elgie R, Cesana BM. GAMIAN–Europe/BEAM survey II: cross–national analysis of unemployment, family history, treatment satisfaction and impact of the bipolar disorder on life style. Bipolar disorders 2004;6(6):487-97.

(5) Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Richardson G, et al. The effectiveness and cost effectiveness of a national lay-led self care support programme for patients with long-term conditions: a pragmatic randomised controlled trial. Journal of Epidemiology and Community Health 2007;61(3):254-61.

(6) National Collaborating Centre for Mental Health, National Institute for Clinical Excellence. Bipolar disorder: The management of bipolar disorder in adults, children, and young people in primary and secondary care - CG38. National Institute for Clinical Excellence; 2006.

(7) Veterans Administration and Department of Defense. VA/DoD clinical practice guideline for management of bipolar disorder in adults. Version 2.0 ed. Washington DC: Veterans Administration; 2010.

(8) Colom F, Vieta E. Psychoeducation manual for bipolar disorder. Cambridge University Press; 2006.

(9) Colom F, Vieta E, Sanchez-Moreno J, Palomino-Otiniano R, Reinares M, Goikolea JM, et al. Group psychoeducation for stabilised bipolar disorders: 5-year outcome of a randomised clinical trial. The British Journal of Psychiatry 2009;194(3):260-5.

(10) Colom F, Vieta E, Martinez-Aran A, Reinares M, Goikolea JM, Benabarre A, et al. A randomized trial on the efficacy of group psychoeducation in the prophylaxis of recurrences in bipolar patients whose disease is in remission. Archives of general psychiatry 2003;60(4):402-7.

(11) Colom F, Vieta E, Reinares M, Martinez-Aran A, Torrent C, Goikolea JM, et al. Psychoeducation efficacy in bipolar disorders: beyond compliance enhancement. The Journal of clinical psychiatry 2003;64(9):1101.

(12) Bond K, Anderson IM. Psychoeducation for relapse prevention in bipolar disorder: a systematic review of efficacy in randomized controlled trials. Bipolar disorders 2015.

(13) Dunn C, Makin N. Clinical Audit Report: Bipolar Disorder (CG185). Lancashire Care Foundation Trust; 2015.

(14) Morriss R, Lobban F, Riste L, Davies L, Holland F, Long R, et al. Clinical effectiveness and acceptability of structured group psychoeducation versus optimised unstructured peer support for patients with remitted bipolar disorder (PARADES): a pragmatic, multicentre, observer-blind, randomised controlled superiority trial. The Lancet Psychiatry 2016.

(15) Morriss RK, Lobban F, Jones S, Riste L, Peters S, Roberts C, et al. Pragmatic randomised controlled trial of group psychoeducation versus group support in the maintenance of bipolar disorder. BMC psychiatry 2011;11(1):114.

(16) Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: Results from a general population study. Health economics 1996;5(2):141-54.

(17) The EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. Health policy 1990;16(3):199-208.

(18) National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal. National Institute for Clinical Excellence; 2013.

(19) Davies LM, Lewis S, Jones PB, Barnes TR, Gaughran F, Hayhurst K, et al. Cost-effectiveness of first-v. second-generation antipsychotic drugs: results from a randomised controlled trial in schizophrenia responding poorly to previous therapy. The British journal of psychiatry: the journal of mental science 2007;191:14-22.

(20) Davies LM, Barnes TR, Jones PB, Lewis S, Gaughran F, Hayhurst K, et al. A Randomized Controlled Trial of the Cost-Utility of Second–Generation Antipsychotics in People with Psychosis and Eligible for Clozapine. Value in Health 2008;11(4):549-62.

(21) Hayhurst H, Palmer S, Abbott R, Johnson T, Scott J. Measuring health-related quality of life in bipolar disorder: relationship of the EuroQol (EQ-5D) to condition-specific measures. Quality of life research 2006;15(7):1271-80.

(22) Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, Donald-Scott P, et al. The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. Archives of general psychiatry 1987;44(6):540-8.

(23) Scott JAN, Paykel E, Morriss R, Bentall R, Kinderman P, Johnson T, et al. Cognitive-behavioural therapy for severe and recurrent bipolar disorders. The British Journal of Psychiatry 2006;188(4):313-20.

(24) Lobban F, Taylor L, Chandler C, Tyler E, Kinderman P, Kolamunnage-Dona R, et al. Enhanced relapse prevention for bipolar disorder by community mental health teams: cluster feasibility randomised trial. The British Journal of Psychiatry 2010;196(1):59-63.

(25) Curtis L. The Unit Costs of Health and Social Care 2012. Kent: Personal Social Services Research Unit; 2013.

(26) Chisholm D, Knapp MRJ, Knudsen HC, Amaddeo F, Gaite LUIS, Van Wijngaarden BOB. Client Socio-Demographic and Service Receipt Inventory-European Version: development of an instrument for international research. The British Journal of Psychiatry 2000;177(39):s28-s33.

(27) Beecham J, Knapp M. Costing psychiatric interventions. Gaskell In Measuring Mental Health Needs. Edited by Thornicroft G. London; 2001. p. 163-83.

(28) Department of Health. National Schedule of Reference Costs 2012-2013. Department of Health; 2013.

(29) Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014;32(12):1157-70.

(30) Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes: Oxford University Press. New York 2005.

(31) Briggs MA, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998;13(4):397-409.

(32) Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science 2007;8(3):206-13.

(33) White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine 2011;30(4):377-99.