ORIGINAL RESEARCH ARTICLE

Using Matching-Adjusted Indirect Comparisons and Network Meta-analyses to Compare Efficacy of Brexanolone Injection with Selective Serotonin Reuptake Inhibitors for Treating Postpartum Depression

Miranda C. Cooper[1] · Hannah S. Kilvert[2] · Paul Hodgkins[3] · Neil S. Roskell[2] · Adi Eldar-Lissai[3]

Abstract

Background Brexanolone injection, the first therapy approved by the US FDA for the treatment of postpartum depression (PPD) in adults, has been shown to produce a significantly greater decrease in the Hamilton Rating Scale for Depression (HAM-D) total score than placebo in randomised controlled trials (RCTs) of women with PPD.

Objectives Given the rapid effect of brexanolone injection (within 60 h) sustained throughout the length of the trials (30 days), we sought to compare its efficacy data against selective serotonin reuptake inhibitors (SSRIs), the class of antidepressants most commonly prescribed for PPD, using HAM-D and Edinburgh Postnatal Depression Scale (EPDS) outcomes from currently available RCTs.

Methods We extracted data from 26 studies identified in a systematic literature review of pharmacological and pharmacological/nonpharmacological combination therapies in PPD. Six studies were suitable to form evidence networks through which to perform indirect treatment comparisons (ITCs) of HAM-D and EPDS outcomes between brexanolone and SSRIs. Having assessed the comparability and suitability of the available evidence for analysis, we discovered significant heterogeneity in the study designs, most notably in the placebo arms of the trials. We therefore conducted matching-adjusted indirect comparisons (MAICs) between brexanolone and the placebo arms of comparator studies, subsequently using the MAIC results of brexanolone versus placebo, and results for SSRIs versus placebo, to form ITCs of brexanolone versus SSRIs at three separate time points—day 3, week 4 and last observation. ITCs were calculated as the differences in change from baseline (CFB) in HAM-D and, separately, CFB in EPDS, between treatments, and reported with 95% confidence intervals (CIs).

Results For all time points, MAICs showed larger differences in CFB for brexanolone compared with SSRIs. Differences (95% CIs) between brexanolone and SSRIs were 12.79 (8.04–17.53) [day 3], 5.87 (− 1.62 to 13.37) [week 4] and 0.97 (− 6.35 to 8.30) [last observation] for the HAM-D. For the EPDS, the differences in CFB were 7.98 (5.32–10.64) [day 3], 6.35 (3.13–9.57) [week 4] and 4.05 (0.79–7.31) [last observation]. Other analytical approaches are also presented to demonstrate the similarity of results, using a network meta-analysis approach, and the importance of using the MAIC method to control for the important heterogeneity between placebo arms.

Conclusions Acknowledging the limitations of ITCs and this evidence base, when compared with SSRIs, these analyses suggest that brexanolone demonstrated larger differences in CFB for both patient- and clinician-reported PPD outcomes and at all investigated time points after adjusting for differences between placebos in the included studies.

✉ Adi Eldar-Lissai
Adi.Eldar-Lissai@sagerx.com

[1] BresMed Health Solutions Ltd, Sheffield, UK

[2] BresMed Health Solutions Ltd, Manchester, UK

[3] Sage Therapeutics, Inc., Cambridge, MA, USA

1 Introduction

Postpartum depression (PPD) is a debilitating mood disorder and one of the most commonly occurring complications in pregnancy and childbirth [1, 2]. The prevalence of PPD across the United States (US) is estimated to be an average of 11.5% (8–20% across states) [3–5], with global estimates of 17.7% (3–38% across 56 countries) [6]. The *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*,

Key Points

Given the differences in study design, matching-adjusted indirect comparisons (MAICs) are needed to compare brexanolone with selective serotonin reuptake inhibitors (SSRIs), as a simple network meta-analysis can give misleading results; key differences include study length and lack of comparability of placebo across studies.

Brexanolone demonstrated a more rapid improvement in postpartum depression compared with SSRIs, as illustrated by the MAIC results at day 3 and week 4.

Brexanolone demonstrated larger differences in change from baseline for both patient- and clinician-reported scales.

defines PPD as a major depressive episode with onset of symptoms occurring during pregnancy or in the 4 weeks following delivery [7], while the American College of Obstetricians and Gynecologists considers PPD to be depression occurring up to 12 months post-delivery [8]. Women with PPD often present with low mood, anxiety, and feelings of depersonalisation and loss of identity [1, 7]. PPD is also associated with an increased risk of substance abuse [9] and, left untreated, can lead to suicidality [10], with suicide itself being a leading cause of maternal mortality [11–14]. Depressive symptoms can persist for up to 11 years following childbirth [15], with up to 30% of women with PPD meeting the criteria of depression during and after the first maternal year [16]. The negative impact of PPD can extend to families, and may lead to partner stress [17] and depression [18], impaired mother–infant bonding [19], problems breastfeeding [20, 21], and isolation from family and friends [1]. Children of mothers with untreated symptoms of PPD are at greater risk of impaired and/or delayed long-term physical, cognitive and psychological outcomes [15, 22–30].

Treatments used to manage PPD include antidepressants, adjunctive to, or in place of, psychotherapy [31–33]. Most studies of pharmacological treatments in PPD focus on selective serotonin reuptake inhibitors (SSRIs), a class of antidepressants often used in the US during the postpartum period despite not being indicated specifically for PPD [31, 34]. Current pharmacological treatments, including SSRIs, are associated with a slow rate of improvement (6–12 weeks) [35], and there a lack of reliable data to support the onset of efficacy occurring within the first week [36, 37]. A Cochrane systematic review of antidepressant treatment in postnatal depression reported that the evidence to support the use of SSRIs in PPD is limited; the few studies published on antidepressants in PPD are generally of low quality, with small sample sizes and risk of bias due to high patient dropout and selective reporting [32]. Obstacles to optimal use of antidepressants include patients receiving subtherapeutic dosing [38], poor compliance [39], and inadequate follow-up due to lack of provider time [40, 41].

A soluble, β-cyclodextrin-based form of the neuroactive steroid allopregnanolone, brexanolone injection (ZULRESSO™) is the first therapy to be approved by the US Food and Drug Administration for the treatment of PPD in adults. In a set of studies in adult women with PPD (Study A: NCT02614547; Study B: NCT02942004; Study C: NCT02942017) [42, 43], brexanolone was associated with a greater change from baseline (CFB) than placebo in the primary outcome measure, the Hamilton Rating Scale for Depression (HAM-D) total score (60-h p value < 0.0001; 30-day p value = 0.0213). A greater CFB for brexanolone was also observed with the Edinburgh Postnatal Depression Scale (EPDS) compared with placebo. However, the EPDS is self-rated and is consequently a secondary outcome measure; therefore, these studies were not powered to show statistically significant differences between brexanolone and placebo on this measure of efficacy (60-h p value = 0.1244; 30-day p value = 0.2515; data on file). The studies showed symptom reduction at 24 h and continued response at 4 weeks.

Given the benefits of brexanolone in PPD demonstrated in these studies, in particular the rapid response sustained throughout the study period, it is important to assess the comparative effectiveness of brexanolone against the current standard of care; such an assessment can help doctors, patients, and formulary decision makers understand the relative efficacy of these treatments. As no head-to-head randomised controlled trial (RCT) comparisons exist, indirect treatment comparison (ITC) approaches are required [44], involving networks of studies linked by one or more common comparators. The suitability of such approaches for use in health technology assessment is well documented [45–47].

We performed a set of ITCs between brexanolone and comparator treatments in PPD, using available RCT data. For this analysis, the HAM-D was selected as an outcome as it is often recognised as the gold-standard outcome measure in studies of depression [48–51], while the EPDS was chosen as it is more commonly used to screen for PPD in clinical practice and to measure symptom burden [32, 49]. Although the EPDS is self-rated, this measure has been shown to be significantly correlated with HAM-D scores within the three brexanolone trials; comparison of treatments based on EPDS is provided as a relevant supporting analysis [52]. In this study, we describe (1) assessing the suitability of identified evidence for comparative efficacy methods, and (2) developing study networks to perform ITCs in order to generate estimates of comparative effectiveness. We discuss our

investigation into the representativeness of the placebo arms of the three brexanolone trials—which have been pooled and are referred to collectively as 547-PPD-202 in this study—and the subsequent development of an appropriate strategy based on matching-adjusted indirect comparison (MAIC) methodology to perform more appropriate comparisons [53].

2 Methods

2.1 Clinical Trials of Brexanolone in Postpartum Depression (PPD)

Brexanolone was investigated in three multicentre, randomised, double-blind, parallel-group, placebo-controlled studies: one phase II trial (NCT02614547; referred to as Study A) and two phase III trials (NCT02942004 and NCT02942017; referred to as Studies B and C, respectively) [42, 43]. In Study B, patients were randomised in a 1:1:1 ratio to brexanolone 90 µg/kg/h (BRX90), brexanolone 60 µg/kg/h (BRX60) or placebo [43], while Studies A and C involved patients randomised in a 1:1 ratio to BRX90 or placebo [42, 43]. Patients in Studies A and B had severe PPD, defined as a HAM-D total score ≥ 26 [42, 43], whereas patients in Study C had moderate PPD (HAM-D score ≥ 20 but ≤ 25) [43]. In all three trials, brexanolone was administered continuously through intravenous infusion, with the dose titrated up to and down from the maximum dose (beginning and end dose, 30 µg/kg/h) [42, 43]. An umbrella protocol facilitated a preplanned analysis of multiple measures of depressive symptoms in an integrated dataset of all three trials [43].

Of note, the placebo arm in 547-PPD-202 differed from traditional placebo. Patients in this arm received medically supervised care during the 60-h infusion, as well as regular observation across the study period, whereas traditional, oral pharmacological therapy placebo arms more closely resemble the level of care expected in outpatient clinical setting (outside of hospital level of care).

2.2 Identification of Studies

A systematic literature review (SLR) performed by Reinhart et al. summarised the evidence for the efficacy, safety and tolerability of pharmacological therapies and pharmacological/nonpharmacological combination therapies in PPD [54]. Of the 32 studies that met the inclusion criteria [49, 51, 55–84], six were observational studies [55–60] and one was a meta-analysis [61], and were thus excluded from the analyses. Following a review update, one further study suitable for analysis was included [85]. Data were extracted from 26 studies of potential relevance (27 studies including the addition of the pooled data from 547-PPD-202) [see electronic supplementary Table 1 for a full list of studies]. The

BRX90 data were the primary focus of these analyses as this was the most robust study arm with the largest sample size. BRX90 data from 547-PPD-202 were pooled together as a single study. The remaining studies were grouped by treatment of interest (brexanolone, SSRIs, and SSRI in combination with nonpharmacological treatments [combination therapies]). The study outcomes of interest were EPDS and HAM-D scores. These outcomes were chosen because while the HAM-D is considered a gold-standard measure in studies of depression [48], the EPDS is more commonly used to screen for PPD in clinical practice and to assess symptom burden [32]. Studies were considered for inclusion in the ITCs if they included at least one pharmacological treatment arm, reported suitable and usable outcome data for the HAM-D and/or EPDS, were randomised (at the patient level), were controlled studies that linked into the ITC network, and were of high enough quality. Figure 1 summarises the study selection process.

2.3 Assessment of Evidence and Development of Strategy

To plan the ITCs, in which all trials included were considered in one cohesive analysis, we assessed the comparability of the studies. The time points used in 547-PPD-202 were considerably different to those used in the other studies; in 547-PPD-202, the EPDS and HAM-D were both measured at 60 h, with follow-up time points through to 30 days, whereas studies of other treatments in the evidence base reported later time points due to longer treatment periods (first assessment at approximately 6–12 weeks). It was therefore considered inappropriate to only compare primary endpoints in the analyses. Time points for CFB were chosen as 60 h (day 3), 30 days (week 4; last time point in 547-PPD-202), and last observation (range across SSRIs studies: 4 weeks–6 months). Missing outcomes were estimated from existing data; for studies where first observation was later than day 3 or week 4, linear interpolation methods were used to impute the missing values.

To investigate the comparability of placebo arms to be used in the analyses, HAM-D and EPDS outcomes were compared for the placebo (control) arms from all studies, for all time points recorded. However, across the comparator studies, placebo arms often included additional care (such as counselling) and therefore may not fully reflect the outcomes of 'traditional' placebo (i.e. neither pharmacological nor nonpharmacological treatment), which should be acknowledged as a limitation. Plots of CFB for EPDS and HAM-D placebo data are given in Figs. 2 and 3, respectively. The improvements in EPDS and HAM-D outcomes in the 547-PPD-202 placebo arm were substantially greater than those of other studies. This difference in magnitude may have led to an underestimation of the relative treatment effect of

Fig. 1 Evidence flowchart.
DE data extraction, *EPDS*
Edinburgh Postnatal Depression
Scale, *GP* general practitioner,
HAM-D Hamilton Rating Scale
for Depression, *PTs* pharmaco-
therapies
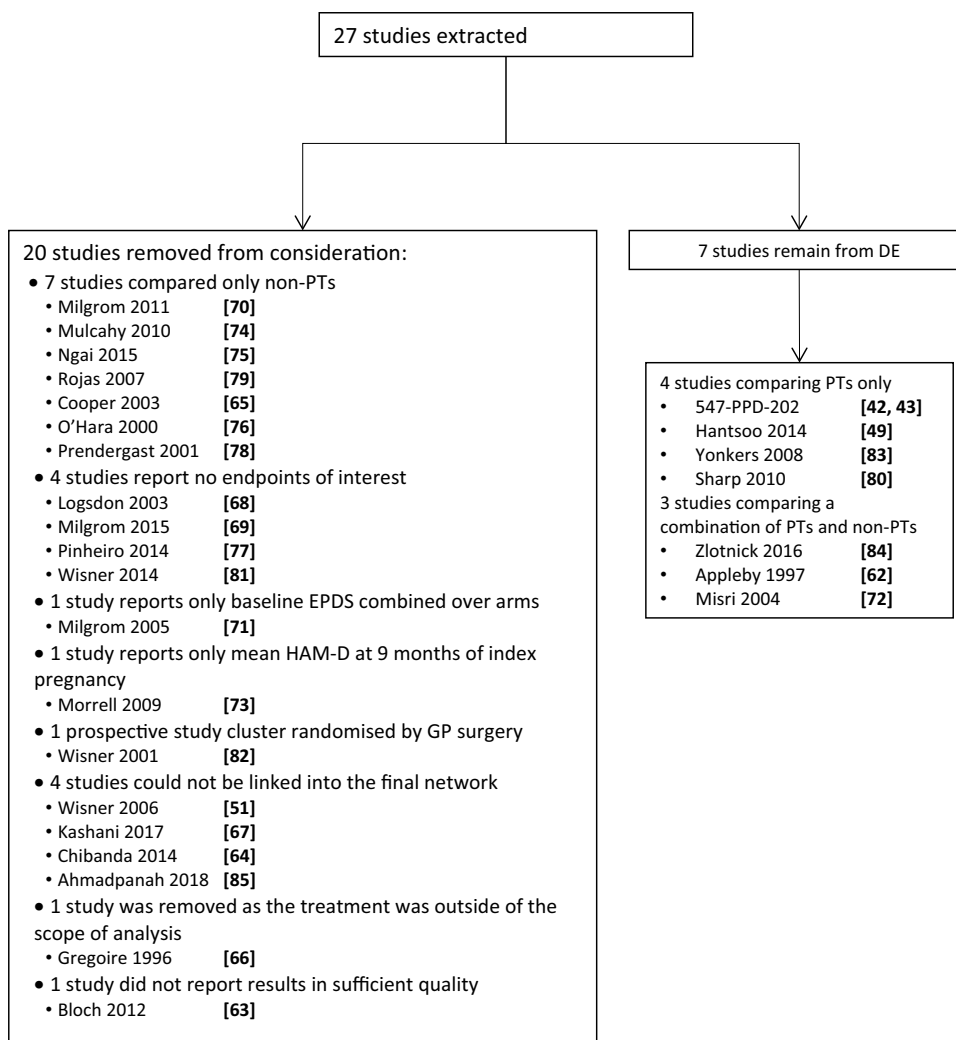
brexanolone against placebo. The time to improvement in the 547-PPD-202 placebo arm was also shorter than that seen in other studies; there was a notable drop in patient- and clinician-reported scores within the first 3 days, followed by a plateau. These data suggest that the 547-PPD-202 placebo arm is not representative of a traditional, oral drug placebo arm and is therefore not suitable as a common comparator in ITCs for brexanolone. As receipt of care and attention in a clinical setting for 3 days is not a viable treatment option, the benefits seen in the 547-PPD-202 placebo arm likely represent an overestimation of the 'effect' of placebo.

2.4 Indirect Treatment Comparison (ITC) Methods

Figure 4 shows the overall ITC strategy.

2.4.1 Bucher ITC

Bucher comparisons are used to compare two treatments that have a common comparator [44]. In the case of two studies,

one comparing treatment A with treatment C and one comparing treatment B with treatment C, treatments A and B can be compared using an indirect link between A and B through treatment C. In this analysis, each indirect estimate was calculated by taking the difference of relative effect values from each study (in this case, the difference in CFB in score, either EPDS or HAM-D, between treatment arms):

$$d_{AB} = d_{AC} - d_{BC},$$

where d_{AB} is the indirect relative effect between treatments A and B, d_{AC} is the relative effect between treatments A and C, and d_{BC} is the relative effect between treatments B and C.

Variance in the indirect estimate was the sum of the variance in the direct estimates:

$$\text{var}(d_{AB}) = \text{var}(d_{AC}) + \text{var}(d_{BC}).$$

2.4.2 Network Meta-analysis

Standard network meta-analysis (NMA) techniques were used in which aggregate relative efficacy estimates for each

Fig. 2 Plot of change from baseline in the Edinburgh Postnatal Depression Scale score reported in the placebo arms [42, 43, 62, 80]. *CFB* change from baseline, *EPDS* Edinburgh Postnatal Depression Scale
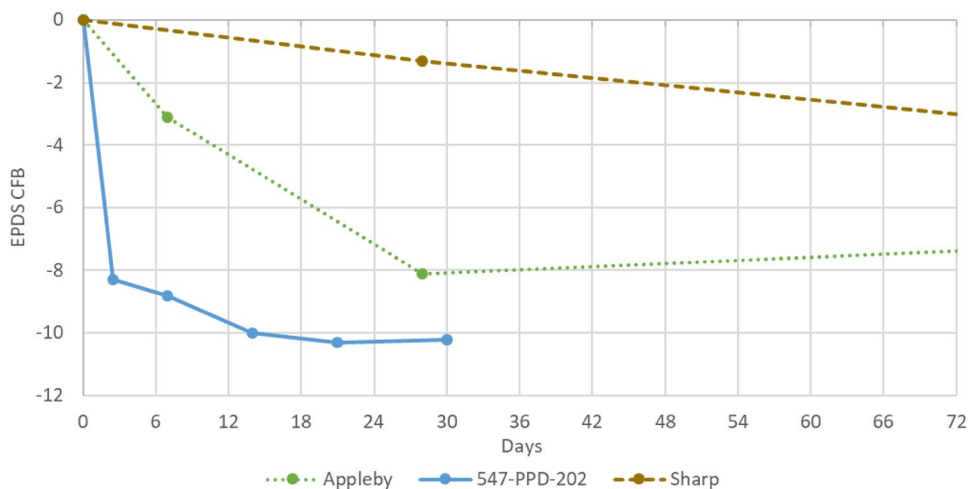

Fig. 3 Plot of change from baseline in the Hamilton Rating Scale for Depression score reported in the placebo arms [42, 43, 49, 62, 83, 84]. *CFB* change from baseline, *HAM-D* Hamilton Rating Scale for Depression
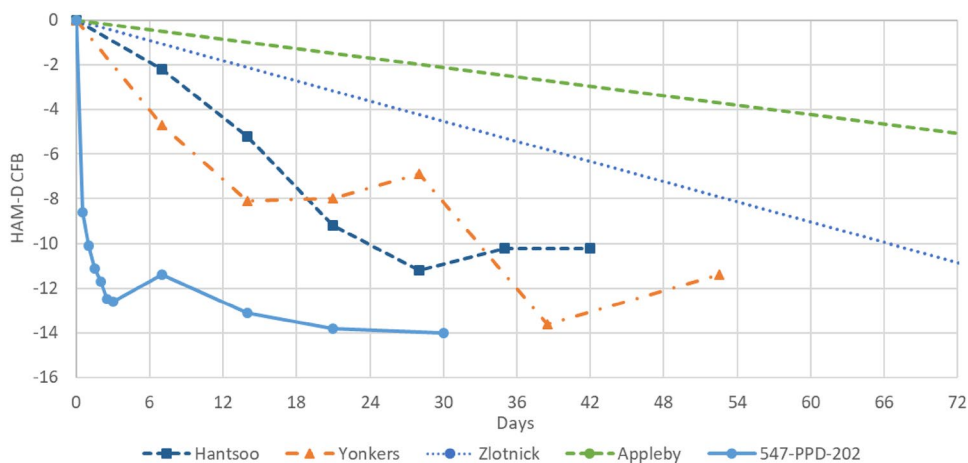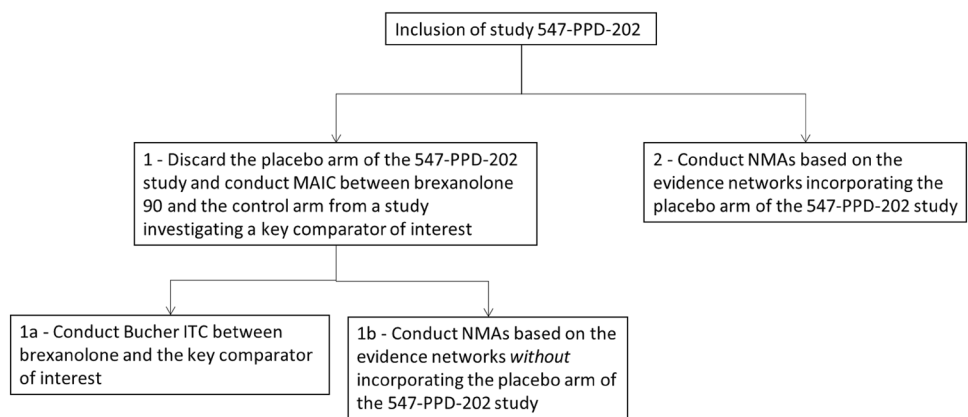

Fig. 4 Flowchart detailing the strategy for estimating the comparative efficacy of brexanolone. *ITC* indirect treatment comparison, *MAIC* matching-adjusted indirect comparison, *NMAs* network meta-analyses

study were incorporated into one consistent analysis under the assumptions of exchangeability [46]. Using the relative effects between the arms of each study (and associated variance), ITCs were made between BRX90 and each of the treatments in the network while preserving randomisation.

As the networks were relatively small, frequentist NMAs were conducted. All models were fitted and run using R software [86], using the 'netmeta' package [87].

2.4.3 Data Availability

The relative effect between treatments was the mean difference in CFB for both outcomes. For CFB measures, the outcome was continuous and was analysed using a normal likelihood. Studies commonly reported a mean and standard deviation (SD) at baseline and post-baseline time points, whereas the ideal outcome measure for use in the ITC analyses would have been CFB (with a measure of variability). The mean CFB, y_{CFB}, was calculated based on the difference between the mean 'before', y_{BL}, and mean 'after', y_{TP}:

$$y_{CFB} = y_{TP} - y_{BL}.$$

To correctly calculate the variance for CFB, V_{CFB}, the variance of baseline values, V_{BL}, the variance of values at the time point, V_{TP}, and the within-patient correlation, ρ, are required:

$$V_{CFB} = V_{TP} + V_{BL} - 2\rho\sqrt{V_{TP}V_{BL}}.$$

However, the within-patient correlation was not available for the studies identified in the SLR. To estimate the variance of the CFB, given that all information was available for 547-PPD-202, the observed variances for CFB were pooled at each time point; these were used as proxies to impute the variance for all studies.

2.4.4 Matching-Adjusted Indirect Comparison

Due to the notable differences between the 547-PPD-202 placebo arm and other studies, the 547-PPD-202 placebo arm was considered inappropriate for forming indirect comparisons. The BRX90 arm was thus treated as a single-arm study, meaning there was no longer a common comparator to connect the BRX90 arm into the evidence network. MAIC methods can allow treatments to be compared indirectly in the absence of a common comparator [53]. In this study, MAICs were conducted between BRX90 and the placebo arm of a chosen comparator study to provide estimated relative effects for BRX90 versus placebo. As matching to placebo was to be performed to facilitate a link into the centre of the network, the evidence base was assessed to choose an appropriate comparator study for the MAICs. For the EPDS-based MAIC, the control arm from the study by Sharp et al. [80] was chosen as although it consisted of listening visits, its large sample size and five variables to match against were a clear advantage. The other option (Appleby et al. [62]) had a smaller sample size and fewer variables to match with, and also included counselling sessions as part of the placebo arm. For the HAM-D-based MAIC, the study by Yonkers et al. [83] was chosen as the comparator study, as while both that study and the study by Zlotnick [84] had larger sample sizes than the other studies and the same number of variables for matching, the control arm in the Zlotnick study involved

clinical management and mothercrafting, making the placebo arm in the study by Yonkers et al. more suitable.

After choosing suitable comparator studies, matching was performed between BRX90 and each chosen control arm by assigning statistical weights to the individual BRX90-treated patients to adjust for their over-/under-representation relative to the comparator source; this was possible due to the availability of 547-PPD-202 patient-level data. In the study by Sharp et al. [80], the patient characteristics used for matching were age of the mother (average and SD), race, antidepressant use at baseline, time since delivery (in weeks) at the start of treatment (average and SD), and whether the mother was primiparous; and in the study by Yonkers et al. [83], the patient characteristics used for matching were age of the mother (average and SD), race, and antidepressant use at baseline. Given the limited data availability, matching was performed on as many variables as possible.

Following weighting, average baseline characteristics (mean and variance) were balanced between the BRX90 and comparator arms. Weights were derived using the propensity score weighting approach proposed by Signorovitch et al. [88, 89]. The propensity score logistic regression model estimated the odds of being enrolled into the BRX90 or comparator arms. The approach uses a method of moments to allow a propensity score logistic regression model to be estimated without comparator study patient-level data. The model was estimated based on the patient-level data available for BRX90-treated patients and the published summary data available for comparator-treated patients.

CFB outcomes for brexanolone were weighted to give weighted mean CFBs, which were incorporated into the ITCs using EPDS and HAM-D data. As inputting into ITCs required associated standard error (SE) values, bootstrap estimators were used to generate SE values and confidence intervals (CIs) for BRX90-weighted mean CFB outcomes to account for the within-subject correlation [90, 91]. First, patients in the BRX90 arm were sampled with replacement to produce a number of bootstrap datasets. For each bootstrap dataset, a set of weights was then derived using the Signorovitch methodology, and, finally, a weighted mean CFB was calculated. This procedure was repeated 1000 times to obtain a distribution of means, and the 2.5th and 97.5th percentiles were used to generate limits of each CI and an SE.

3 Results

3.1 Identification of Evidence in PPD

Following data extraction and the exclusion of studies (Sect. 2.2), the final evidence base comprised six studies, i.e. Appleby et al. [62], Sharp et al. [80], Misri et al. [72],

Hantsoo et al. [49], Yonkers et al. [83], and Zlotnick [84]. The clinical heterogeneity of the evidence base was assessed by comparing baseline patient and study characteristics (see electronic supplementary Table 2). While some aspects differed across the trials, in most cases the differences were deemed insufficiently clear or not reported well enough to lead to clear biases in outcomes. We therefore did not make additional adjustments or perform sensitivity analyses based on baseline characteristics.

Of the remaining studies, three reported EPDS outcomes (Appleby et al. [62], Sharp et al. [80] and Misri et al. [72]) and five reported HAM-D outcomes (Appleby et al. [62], Hantsoo et al. [49], Yonkers et al. [83], Zlotnick [84], and Misri et al. [72]). These studies were used to establish network diagrams for the EPDS and HAM-D (as shown in Figs. 5 and 6, respectively). To reflect clinical practice, in which therapies are often considered by treatment class rather than individually, comparator treatments were grouped into two categories—SSRIs, and combination therapies. To construct a connected evidence network, equivalence was assumed for the placebo arms of the included studies.

3.2 ITC Strategy

Given the fundamental incompatibility of 547-PPD-202 placebo data with those of comparator studies, we developed two approaches to comparing BRX90 with comparators in order to assess the impact of including 547-PPD-202 placebo data (represented in Fig. 4):
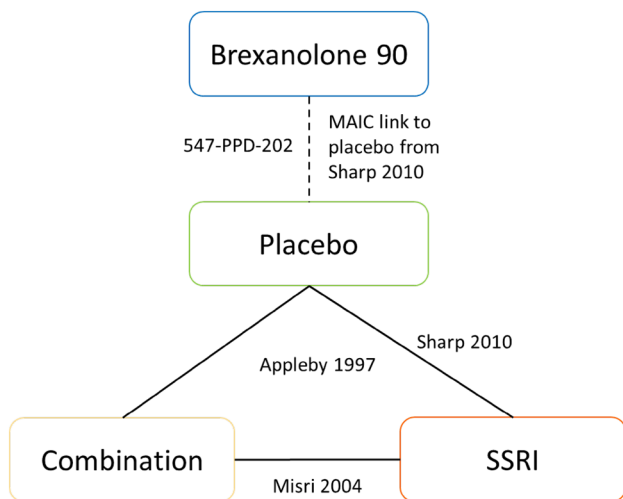

Fig. 5 Evidence network—Edinburgh Postnatal Depression Scale. For the unadjusted network meta-analyses, brexanolone is linked through the 547-PPD-202 placebo arm [42, 43, 62, 72, 80]. *MAIC* matching-adjusted indirect comparison, *SSRI* selective serotonin reuptake inhibitor
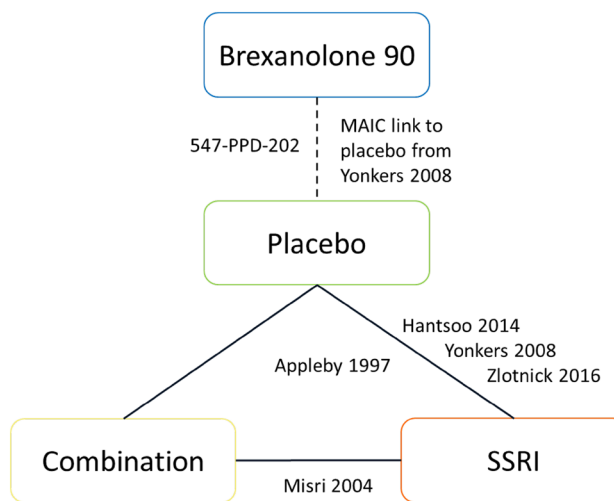

Fig. 6 Evidence network—Hamilton Rating Scale for Depression. For the unadjusted network meta-analyses, brexanolone is linked through the 547-PPD-202 placebo arm [42, 43, 49, 62, 72, 83, 84]. *MAIC* matching-adjusted indirect comparison, *SSRI* selective serotonin reuptake inhibitor

- conduct a set of ITCs using the relative treatment effects from 547-PPD-202;
- conduct a set of analyses in which the 547-PPD-202 placebo data are discarded, effectively treating 547-PPD-202 as a BRX90 single-arm trial.

This latter approach involved using the available 547-PPD-202 patient-level data to perform MAICs between BRX90 and the placebo arms of the chosen comparator studies, giving estimated relative efficacy outcomes. The weighted mean BRX90 outcomes were then linked into evidence networks as part of the ITCs.

We present the results of unadjusted ITCs (i.e. Bucher ITCs and NMAs conducted without MAIC) and adjusted ITCs (i.e. Bucher ITCs and NMAs using the MAIC-estimated relative treatment effects of BRX90, to compensate for treating 547-PPD-202 as a single-arm trial) between BRX90 and SSRIs. A diagram of the overall strategy is given in Fig. 4. The two studies used in matching, i.e. Sharp et al. [80] (EPDS) and Yonkers et al. [83] (HAM-D), were used in Bucher ITCs to evaluate the effect of matching versus including the 547-PPD-202 placebo. Comparisons with combination therapies and placebo are presented in the electronic supplementary material.

3.3 Unadjusted and Adjusted ITC Analyses

Forest plots of ITCs for BRX90 versus SSRIs are presented in Fig. 7 (EPDS) and Fig. 8 (HAM-D). The results of all ITC analyses of BRX90 versus SSRIs, combination
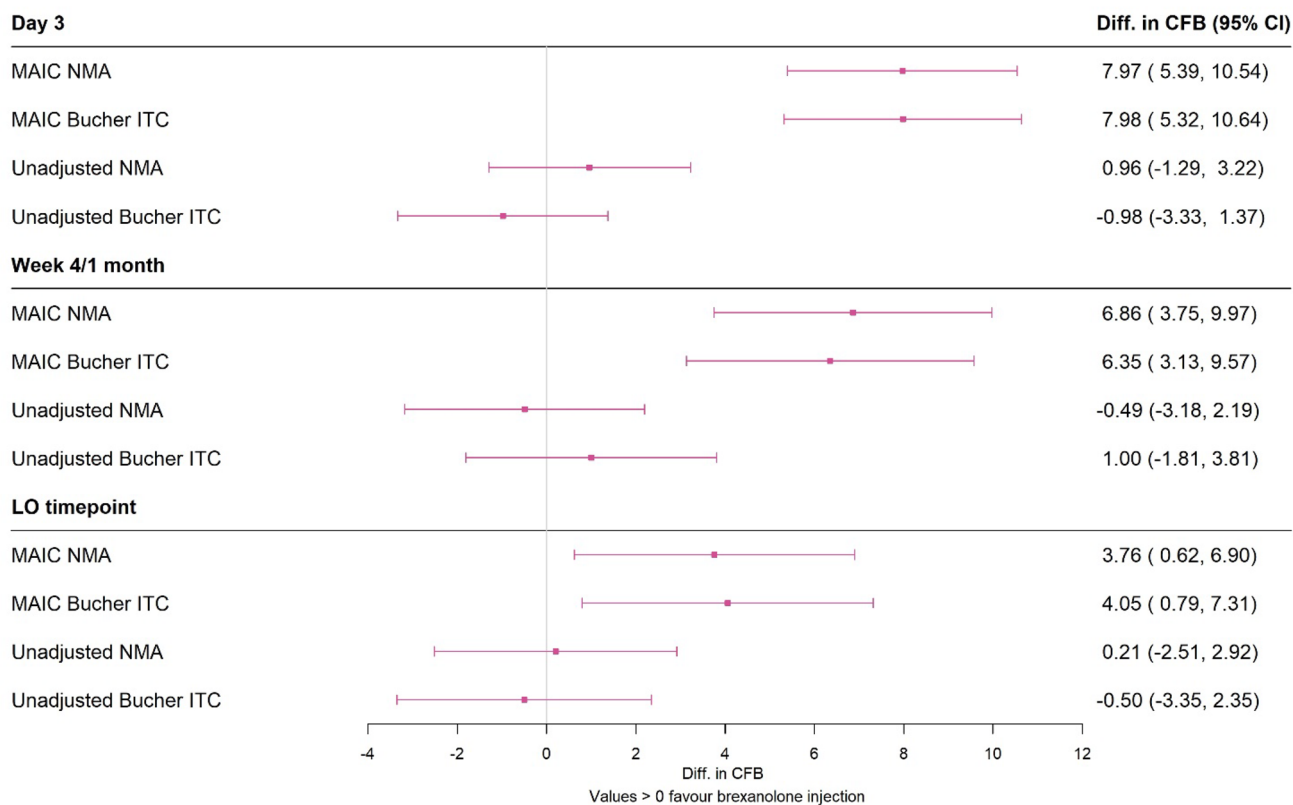
Fig. 7 Brexanolone versus selective serotonin reuptake inhibitors: unadjusted and adjusted indirect treatment comparison results—Edinburgh Postnatal Depression Scale. Bucher ITC and NMAs include different studies. LO time point ranged from 4 to 18 weeks. *CI* con-fidence interval, *CFB* change from baseline, *ITC* indirect treatment comparison, *LO* last observation, *MAIC* matching-adjusted indirect comparison, *NMA* network meta-analysis, *Diff.* Difference

therapies, and placebo are presented in electronic supplementary Table 3 (EPDS) and electronic supplementary Table 4 (HAM-D). The full results of the MAICs performed between BRX90 and the placebo arms of the chosen comparator studies, i.e. Sharp et al. [80] (EPDS) and Yonkers et al. [83] (HAM-D) are given in electronic supplementary Tables 5 and 6.

In both the MAIC-adjusted Bucher ITCs and NMAs for day 3, the CFB in the HAM-D and EPDS for BRX90 was greater than that for SSRIs (HAM-D: 12.79, 95% CI 8.04–17.53 [Bucher ITC] and 13.47, 95% CI 9.84–17.10 [NMA]; EPDS: 7.98, 95% CI 5.32–10.64 [Bucher ITC] and 7.97, 95% CI 5.39–10.54) [NMA]). Based on the CIs, none of which crossed zero, these differences were indicated to be significant. In comparison, unadjusted Bucher ITCs and NMAs for day 3 did not suggest significant differences between BRX90 and SSRIs for the EPDS, whereas for the HAM-D, significance was maintained for the NMA SSRI comparison. For both the MAIC and unadjusted analyses, results for the comparison with SSRIs were generally consistent between the two methods considered (Bucher ITC and NMA).

At week 4, SSRIs were indicated to have significantly smaller reductions in EPDS compared with BRX90, in both the MAIC-adjusted Bucher ITC and NMA (Bucher ITC: 6.35, 95% CI 3.13–9.57; NMA: 6.86, 95% CI 3.75–9.97). Similarly, in the week 4 MAIC-adjusted Bucher ITC for the HAM-D, SSRIs continued to lead to a smaller reduction in HAM-D compared with BRX90, albeit one that was indicated as nonsignificant (5.87, 95% CI −1.62 to 13.37). However, the MAIC-adjusted NMA for the HAM-D, at the same time point, gave a significantly larger difference for BRX90 compared with SSRIs (8.41, 95% CI 2.94–13.88). The differences in the EPDS and HAM-D CFB between BRX90 and SSRIs had smaller absolute values than for day 3, likely due to the extra time available for SSRI efficacy to increase. In comparison, the unadjusted analyses for week 4 showed no significant differences between BRX90 and SSRIs.

In the last observation time point (EPDS: weeks 4–18; HAM-D: week 4 to 6 months), when SSRIs were assumed to have been able to reach maximum efficacy, the MAIC-adjusted Bucher ITC and NMA indicated that SSRIs still had smaller reductions in the EPDS and HAM-D compared with BRX90 (HAM-D: 0.97, 95% CI −6.35 to 8.30 [Bucher ITC] and 2.02, 95% CI −3.18 to 7.22 [NMA]; EPDS: 4.05,
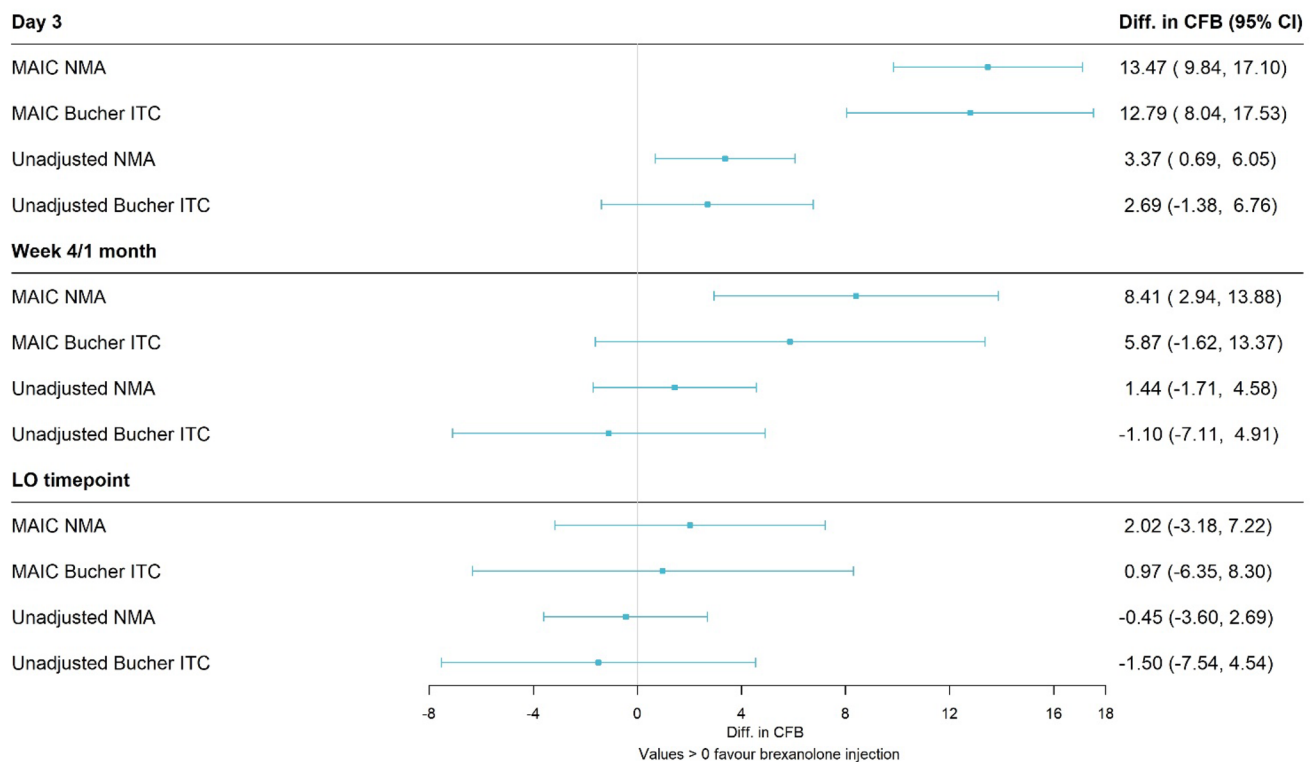
Fig. 8 Brexanolone versus selective serotonin reuptake inhibitors: unadjusted and adjusted indirect treatment comparison results—Hamilton Rating Scale for Depression. Bucher ITC and NMAs include different studies. LO time point ranged from 4 weeks to 6 months. *CI* confidence interval, *CFB* change from baseline, *ITC* indirect treatment comparison, *LO* last observation, *MAIC* matching-adjusted indirect comparison, *NMA* network meta-analysis, *Diff.* difference

95% CI 0.79–7.31 [Bucher ITC] and 3.76, 95% CI 0.62–6.90 [NMA]). However, at this time point, the magnitude of the differences between BRX90 and SSRIs were smallest and the differences for the HAM-D were no longer significant. Similar to previous time points, the unadjusted analyses at the last observation time point showed nonsignificant differences in CFB between BRX90 and SSRIs.

The results for the MAIC-adjusted analyses demonstrated that at all time points, BRX90 led to a greater CFB than SSRIs, for both EPDS and HAM-D. In contrast, the unadjusted analyses showed little difference between BRX90 and SSRIs. At each time point, the key differences in results lie between the MAIC-adjusted and -unadjusted analyses. The choice of analytical approach is therefore crucial to understanding and accurately reflecting the benefit of BRX90.

4 Discussion

As highlighted by Meltzer-Brody et al. [43], treatments for PPD are urgently needed as there is an increased risk of morbidity and mortality in mothers with PPD [92]. The effects of brexanolone on children and families of patients were not investigated in 547-PPD-202; however, the rapid and sustained symptom reduction is potentially beneficial for patients and families alike. Netsi et al. investigated the long-term effects of PPD within a large observational study of women and their offspring and found that children of women with persistent and severe PPD showed an increased risk for behavioural problems by age 3.5 years, along with lower mathematics grades and higher prevalence of depression during adolescence. Furthermore, mothers with chronic PPD could suffer from depressive symptoms 11 years after childbirth [15]. Other studies have shown an association of maternal depression with increased infanticide and decreased use of preventive paediatric health care [93, 94]. Therefore, the rapid and sustained symptom reduction shown in the 547-PPD-202 study, along with greater efficacy when compared with SSRIs via the ITC, can potentially lead to better treatment and symptom control for mothers with PPD.

In this study, we investigated the current clinical evidence regarding the efficacy and speed of improvement for treatments of PPD, conducting comparative efficacy analysis of brexanolone with SSRIs. In the absence of head-to-head trials, we assessed the comparability of identified RCTs and performed ITCs of treatment efficacy. Overall, the MAIC-adjusted ITC results suggest a greater efficacy of BRX90 compared with SSRIs, for both patient-reported

(EPDS) and clinician-reported (HAM-D) outcomes and at all time points measured. In the MAIC-adjusted Bucher ITC and NMA at the last observation time point (EPDS: weeks 4–18; HAM-D: week 4 to 6 months), despite all treatments being assumed to have had the opportunity to reach optimum efficacy, BRX90 still gave larger CFB than SSRIs, albeit with smallest absolute values, compared with treatment differences at earlier time points. Thus, brexanolone efficacy is comparatively rapid and also shows sustained relative efficacy compared with these comparator groups during the study period.

As noted in a recent Cochrane review for antidepressants in PPD, the evidence base for this analysis comprised a few low-quality studies with small sample sizes, many of which were performed approximately a decade ago [32]. The severity of depression at baseline varied between all trials included in these analyses, and it is not known whether and to what extent this affected the ITC results. Notably, the difference between the greatest and lowest baseline score was larger than the minimally important difference (HAM-D: 2.1; EPDS: 2.9; data on file), indicating that choosing a different placebo arm for matching may skew the results. Specifically, the EPDS score at baseline of the pooled 547-PPD-202 BRX90 arm was greater than that of the placebo arm in the study by Sharp et al. [80], whereas for the placebo arm in the study by Yonkers et al. [83] and 547-PPD-202, the HAM-D scores were similar at baseline. Given the evidence base, each of these studies was considered best suited for matching. While there is a larger evidence base for major depressive disorder (MDD), we considered its use in this study inappropriate as patient populations of studies in MDD and PPD are not compatible with respect to sex and age range.

The placebo arms of the comparator studies differed in terms of study design. Some placebo arms also included behavioural therapy interventions; for example, the study by Sharp et al. [80] included listening visits, and the study by Appleby et al. [62] included one counselling session. To construct study networks for NMAs, the placebo arms of comparator studies were assumed to be equivalent, whereas many of these involved nonpharmacological interventions. Similarly, the trials reporting data for SSRIs were grouped, which assumed equivalent efficacy of SSRIs. As Appleby et al. [62] reported geometric mean values, these were assumed to be comparable with arithmetic means reported by other studies, to allow treatment comparisons to be made, which may not be appropriate.

The outcomes of two measures were used in the analysis—HAM-D and EPDS. While the EPDS was developed as a screening tool, both in clinical practice and clinical trials, it was used to measure the extent of symptoms of PPD; therefore, this tool may not have captured all relevant aspects of changing PPD symptoms as it was not built for this purpose.

One of the key clinical advantages of brexanolone, i.e. the rapid response seen in patients, leads to issues with performing ITCs against studies of treatments with slower onset of efficacy. 547-PPD-202 had approximately 4 weeks follow-up post-treatment, but its study period of 30 days was shorter than the SSRI studies, which ranged from 4 weeks to 6 months. To impute the EPDS and HAM-D for data missing at earlier time points, linear interpolation methods were used. Extending a linear trend from the data may not represent the true trajectory of comparator treatment outcomes over time. Similarly, the use of last observation carried forward to provide estimates of the missing values for brexanolone at time points later than 30 days assumed that the efficacy of brexanolone was sustained past this final measurement. Furthermore, while the choice of last time point allowed SSRIs to reach their optimal efficacy, the assumption that brexanolone reached optimal efficacy at 30 days prohibited any change after this point. Imputations were also made for studies that did not report the CFB variances, instead using observed data from the 547-PPD-202 study. We cannot know whether these imputations over- or underestimated variances, and we do not expect that these imputations created bias for or against brexanolone.

The use of MAICs was deemed a necessary step in the analyses; however, as with all analytical approaches, the MAIC method has limitations. With the 547-PPD-202 placebo arms discarded, there is no common comparator when conducting ITCs, which would be the most robust method; the *National Institute for Health and Care Excellence Decision Support Unit Technical Support Document 18* acknowledges unanchored analyses as more likely to be biased than anchored analyses [53]. The same guidance document highlights that with unanchored comparisons, all treatment effect modifiers or prognostic factors should be considered. However, it is unlikely that all key study characteristics were captured during matching as only a small number of relevant characteristics were available in both 547-PPD-202 and reported comparator studies. Furthermore, the MAIC-adjusted NMA required the assumption that the estimated relative effects of BRX90 were constant across the patient population to which the BRX90 patient-level data were matched and the patient populations of the rest of the network. In addition, the choice of study placebo arm with which to match/adjust brexanolone data is subjective and different choices may lead to different results. Across any choice of study placebo arm to match to, brexanolone, an intravenous drug, is compared with an oral placebo. As these analyses are retrospective, the impact of the different treatment administrations is not anticipated to strongly influence comparisons of efficacy outcomes. However, the impact of oral versus intravenous administration is recognised to influence the decision-making process when considering PPD treatment pathways. This impact is anticipated to be

reflected through other measures such as quality of life and cost implications. Similarly, payers and decision makers should also consider safety implications associated with the treatment options.

While there are limitations associated with performing ITCs with the identified studies given the difference in endpoint times, the justification for this is to account for the relatively rapid mode of action of brexanolone, a significant clinical and patient-relevant advantage. It is of note that the imputation methods used produce an SSRI effect at day 3; however, it is uncertain whether this is an actual treatment effect as investigations into whether SSRIs have an effect by 3 days are limited and have led to mixed results [36, 37]. Therefore, the interpolation for day 3 results may be biased in favour of SSRIs. However, we recognise that the assumption the efficacy of brexanolone continues beyond the observed trial period may bias the analyses conducted at this time point, although the direction of this bias is unclear. As outcomes at a longer follow-up time for patients treated with brexanolone were not available, this assumption was required in order to make comparisons with SSRIs that better reflect their long-term time-to-efficacy. It is worth noting that the MAIC-estimated treatment differences between brexanolone and SSRIs at both day 3 and week 4 are deemed clinically relevant.

Where there is only one treatment comparison of interest, which can be informed by two studies and a common comparator, the Bucher approach to ITCs may be favoured because of its transparency and its reliance on fewer assumptions (versus a broad NMA with numerous links between studies). However, NMAs have the flexibility to include more treatments and corresponding studies, although they do rely on additional assumptions of homogeneity of the included studies. Across the MAIC-adjusted and unadjusted Bucher ITCs and NMAs, the results of the BRX90 comparisons with SSRIs were broadly similar, for both the EPDS and HAM-D, which may support the use of the simpler method.

5 Conclusions

We have demonstrated the use of MAIC techniques to perform a more appropriate comparison between BRX90 and SSRIs in the absence of a suitable placebo arm. The MAICs suggest that BRX90 provides a significant improvement for the HAM-D and EPDS outcomes at day 3 relative to SSRIs. Despite a decrease in the absolute difference between BRX90 and SSRIs over time for the HAM-D and EPDS outcomes, the MAICs still suggest that at later time points, BRX90 generally provides equivalent or improved outcomes compared with SSRIs.

Despite the limitations associated with MAICs, such an approach was needed to avoid the use of the unrepresentative 547-PPD-202 placebo arm, with a significantly greater magnitude of improvement than placebo arms of those studies compared, and a methodology (around-the-clock monitoring for a 3-day period) that is not a viable option in clinical practice. Otherwise, standard NMA methodology in which these data were used would likely be misleading as to the effect of brexanolone. Thus, the ITCs in which MAICs were used are likely to capture the true incremental efficacy of brexanolone. Given the ITC limitations regarding the lack of consistency of trial design, especially the availability of outcome data at the different time points, the analysis shown for week 4 (1 month) could be deemed more robust than the other analyses. It is of clinical relevance to provide some estimates of relative treatment effects, acknowledging limitations, at a very early time point (day 3) and also at a later time point (where maximum effect is assumed to be realised).

Compliance with Ethical Standards

References

1. Beck CT. Postpartum depression: it isn't just the blues. Am J Nurs. 2006;106(5):40–50.
2. Howard MM, Mehta ND, Powrie R. Peripartum depression: early recognition improves outcomes. Clevel Clin J Med. 2017;84(5):388–96.
3. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol. 2005;106(5 Pt 1):1071–83.
4. Gjerdingen DK, Yawn BP. Postpartum depression screening: importance, methods, barriers, and recommendations for practice. J Am Board Fam Med. 2007;20(3):280–8.

5. Ko JY, Rockhill KM, Tong VT, Morrow B, Farr SL. Trends in postpartum depressive symptoms—27 States, 2004, 2008, and 2012. MMWR Morb Mortal Wkly Rep. 2017;66(6):153–8.

6. Hahn-Holbrook J, Cornwell-Hinrichs T, Anaya I. Economic and health predictors of national postpartum depression prevalence: a systematic review, meta-analysis, and meta-regression of 291 studies from 56 countries. Front Psychiatry. 2017;8:248.

7. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Publishing; 2013.

8. American College of Obstetricians and Gynecologists. Postpartum depression. Patient FAQs (FAQ091). 2013. https://www.acog.org/Patients/FAQs/Postpartum-Depression Accessed 1 Apr 2019.

9. Fitelson E, Kim S, Baker AS, Leight K. Treatment of postpartum depression: clinical, psychological and pharmacological options. Int J Womens Health. 2010;3:1–14.

10. Moses-Kolko EL, Hipwell AE. First-onset postpartum psychiatric disorders portend high 1-year unnatural-cause mortality risk. Am J Psychiatry. 2016;173(6):559–61.

11. Gressier F, Guillard V, Cazas O, Falissard B, Glangeaud-Freudenthal NM, Sutter-Dallay AL. Risk factors for suicide attempt in pregnancy and the post-partum period in women with serious mental illnesses. J Psychiatr Res. 2017;84:284–91.

12. Palladino CL, Singh V, Campbell J, Flynn H, Gold KJ. Homicide and suicide during the perinatal period: findings from the National Violent Death Reporting System. Obstet Gynecol. 2011;118(5):1056–63.

13. Bodnar-Deren S, Klipstein K, Fersh M, Shemesh E, Howell EA. Suicidal ideation during the postpartum period. J Womens Health (2002). 2016;25(12):1219–24.

14. Lindahl V, Pearson JL, Colpe L. Prevalence of suicidality during pregnancy and the postpartum. Arch Womens Ment Health. 2005;8(2):77–87.

15. Netsi E, Pearson RM, Murray L, Cooper P, Craske MG, Stein A. Association of persistent and severe postnatal depression with child outcomes. JAMA Psychiatry. 2018;75(3):247–53.

16. Vliegen N, Casalin S, Luyten P. The course of postpartum depression: a review of longitudinal studies. Harv Rev Psychiatry. 2014;22(1):1–22.

17. Kerstis B, Nohlert E, Öhrvik J, Widarsson M. Association between depressive symptoms and parental stress among mothers and fathers in early parenthood: a Swedish cohort study. Upsala J Med Sci. 2016;121(1):60–4.

18. Goodman JH. Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. J Adv Nurs. 2004;45(1):26–35.

19. Lilja G, Edhborg M, Nissen E. Depressive mood in women at childbirth predicts their mood and relationship with infant and partner during the first year postpartum. Scand J Caring Sci. 2012;26(2):245–53.

20. Gagliardi L, Petrozzi A, Rusconi F. Symptoms of maternal depression immediately after delivery predict unsuccessful breast feeding. Archives of disease in childhood. 2012;97(4):355–7.

21. Wouk K, Lara-Cinisomo S, Stuebe AM, Poole C, Petrick JL, McKenney KM. Clinical interventions to promote breastfeeding by Latinas: a meta-analysis. Pediatrics. 2016;137(1):e20152423.

22. Eastwood JG, Jalaludin BB, Kemp LA, Phung HN, Barnett BE. Relationship of postnatal depressive symptoms to infant temperament, maternal expectations, social support and other potential risk factors: findings from a large Australian cross-sectional study. BMC Pregnancy Childb. 2012;12:148.

23. Woolhouse H, Gartland D, Mensah F, Giallo R, Brown S. Maternal depression from pregnancy to 4 years postpartum and emotional/behavioural difficulties in children: results from a prospective pregnancy cohort study. Arch Womens Ment Health. 2016;19(1):141–51.

24. Verkuijl NE, Richter L, Norris SA, Stein A, Avan B, Ramchandani PG. Postnatal depressive symptoms and child psychological development at 10 years: a prospective study of longitudinal data from the South African Birth to Twenty cohort. Lancet Psychiatry. 2014;1(6):454–60.

25. Pearson RM, Evans J, Kounali D, Lewis G, Heron J, Ramchandani PG, et al. Maternal depression during pregnancy and the postnatal period: risks and possible mechanisms for offspring depression at age 18 years. JAMA Psychiatry. 2013;70(12):1312–9.

26. Kawai E, Takagai S, Takei N, Itoh H, Kanayama N, Tsuchiya K. Maternal postpartum depressive symptoms predict delay in nonverbal communication in 14-month-old infants. Infant Behav Dev. 2017;46:33–45.

27. Yamaoka Y, Fujiwara T, Tamiya N. Association between maternal postpartum depression and unintentional injury among 4-month-old infants in Japan. Matern Child Health J. 2016;20(2):326–36.

28. Koutra K, Chatzi L, Bagkeris M, Vassilaki M, Bitsios P, Kogevinas M. Antenatal and postnatal maternal mental health as determinants of infant neurodevelopment at 18 months of age in a mother–child cohort (Rhea Study) in Crete, Greece. Soc Psychiatry Psychiatr Epidemiol. 2013;48(8):1335–45.

29. Surkan PJ, Ettinger AK, Hock RS, Ahmed S, Strobino DM, Minkovitz CS. Early maternal depressive symptoms and child growth trajectories: a longitudinal analysis of a nationally representative US birth cohort. BMC Pediatr. 2014;14(1):185.

30. Moore Simas TA, Huang MY, Patton C, Reinhart M, Chawla AJ, Clemson C, et al. The humanistic burden of postpartum depression: a systematic literature review. Curr Med Res Opin. 2019;35(3):383–93.

31. Sharma V, Sommerdyk C. Are antidepressants effective in the treatment of postpartum depression? A systematic review. Prim Care Companion CNS Disord. 2013;15(6).

32. Molyneaux E, Howard LM, McGeown HR, Karia AM, Trevillion K. Antidepressant treatment for postnatal depression. Cochrane Database Syst Rev. 2014;9:CD002018.

33. Stewart DE, Vigod S. Postpartum depression. N Engl J Med. 2016;375(22):2177–86.

34. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder: third edition. 2010. https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf. Accessed 11 Mar 2019.

35. Frazer A, Benmansour S. Delayed pharmacological effects of antidepressants. Mol Psychiatry. 2002;7(Suppl 1):S23–8.

36. Gelenberg AJ, Chesen CL. How fast are antidepressants? J Clin Psychiatry. 2000;61(10):712–21.

37. Machado-Vieira R, Salvadore G, Luckenbaugh DA, Manji HK, Zarate CA Jr. Rapid onset of antidepressant action: a new paradigm in the research and treatment of major depressive disorder. J Clin Psychiatry. 2008;69(6):946–58.

38. Bushnell GA, Sturmer T, Swanson SA, White A, Azrael D, Pate V, et al. Dosing of selective serotonin reuptake inhibitors among children and adults before and after the FDA black-box warning. Psychiatr Serv. 2016;67(3):302–9.

39. Lingam R, Scott J. Treatment non-adherence in affective disorders. Acta Psychiatr Scand. 2002;105(3):164–72.

40. Byatt N, Simas TA, Lundquist RS, Johnson JV, Ziedonis DM. Strategies for improving perinatal depression treatment in North American outpatient obstetric settings. J Psychosom Obstet Gynaecol. 2012;33(4):143–61.

41. Jones LE, Turvey C, Carney-Doebbeling C. Inadequate follow-up care for depression and its impact on antidepressant treatment duration among veterans with and without diabetes mellitus in the Veterans Health Administration. Gen Hosp Psychiatry. 2006;28(6):465–74.

42. Kanes S, Colquhoun H, Gunduz-Bruce H, Raines S, Arnold R, Schacterle A, et al. Brexanolone (SAGE-547 injection) in

post-partum depression: a randomised controlled trial. Lancet. 2017;390(10093):480–9.

43. Meltzer-Brody S, Colquhoun H, Riesenberg R, Epperson CN, Deligiannidis KM, Rubinow DR, et al. Brexanolone injection in post-partum depression: two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. Lancet. 2018;392(10152):1058–70.

44. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683–91.

45. Dias S, Welton N, Sutton A, Ades A. NICE DSU Technical support document 1: introduction to evidence synthesis for decision making. 2011. http://www.nicedsu.org.uk. Accessed 13 Apr 2018.

46. Dias S, Welton N, Sutton A, Ades A. NICE DSU Technical support document 2: a general linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011 (last updated Sep 2016). http://www.nicedsu.org.uk. Accessed 13 Apr 2018.

47. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. Value Health. 2011;14(4):429–37.

48. Khan A, Khan SR, Shankles EB, Polissar NL. Relative sensitivity of the Montgomery–Asberg Depression Rating Scale, the Hamilton Depression rating scale and the clinical global impressions rating scale in antidepressant clinical trials. Int Clin Psychopharmacol. 2002;17(6):281–5.

49. Hantsoo L, Ward-O'Brien D, Czarkowski KA, Gueorguieva R, Price LH, Epperson CN. A randomized, placebo-controlled, double-blind trial of sertraline for postpartum depression. Psychopharmacology (Berl). 2014;231(5):939–48.

50. Kim DR, Epperson CN, Weiss AR, Wisner KL. Pharmacotherapy of postpartum depression: an update. Expert Opin Pharmacother. 2014;15(9):1223–34.

51. Wisner KL, Hanusa BH, Perel JM, Peindl KS, Piontek CM, Sit DK, et al. Postpartum depression: a randomized trial of sertraline versus nortriptyline. J Clin Psychopharmacol. 2006;26(4):353–60.

52. Gerbasi ME, Meltzer-Brody S, Eldar-Lissai A, Acaster S, Fridman M, Bonthapally V, et al. The association between the Hamilton Rating Scale for Depression and the Edinburgh Postnatal Depression Scale in postpartum depression. American Psychiatric Association annual meeting, 18–22 May 2019, San Francisco, CA.

53. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical support document 18: methods for population-adjusted indirect comparisons in submission to NICE. 2016. http://www.nicedsu.org.uk. Accessed 13 Apr 2018.

54. Reinhart M, Patton C, Chawla A, Clemson C, Huang M, Bonthapally V. Treatment of postpartum depression: a systematic literature review. Value Health. 2017;20(9):A717.

55. Hendrick V, Altshuler L, Strouse T, Grosser S. Postpartum and nonpostpartum depression: differences in presentation and response to pharmacologic treatment. Depress Anxiety. 2000;11(2):66–72.

56. Logsdon MC, Wisner K, Sit D, Luther JF, Wisniewski SR. Depression treatment and maternal functioning. Depress Anxiety. 2011;28(11):1020–6.

57. Kuosmanen L, Vuorilehto M, Kumpuniemi S, Melartin T. Postnatal depression screening and treatment in maternity and child health clinics. J Psychiatr Mental Health Nurs. 2010;17(6):554–7.

58. Pearlstein TB, Zlotnick C, Battle CL, Stuart S, O'Hara MW, Price AB, et al. Patient choice of treatment for postpartum depression: a pilot study. Arch Womens Ment Health. 2006;9(6):303–8.

59. Highet N, Drummond P. A comparative evaluation of community treatments for post-partum depression: implications for treatment and management practices. Aust N Z J Psychiatry. 2004;38(4):212–8.

60. Fam J, Chen H, Wang J. Supportive counseling for postpartum depression in Asian mothers. Asia Pac Psychiatry. 2011;3(2):61–6.

61. McDonagh MS, Matthews A, Phillipi C, Romm J, Peterson K, Thakurta S, et al. Depression drug treatment outcomes in pregnancy and the postpartum period: a systematic review and meta-analysis. Obstet Gynecol. 2014;124(3):526–34.

62. Appleby L, Warner R, Whitton A, Faragher B. A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. BMJ. 1997;314(7085):932.

63. Bloch M, Meiboom H, Lorberblatt M, Bluvstein I, Aharonov I, Schreiber S. The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of postpartum depression: a randomized, double-blind, placebo-controlled study. J Clin Psychiatry. 2012;73(2):235–41.

64. Chibanda D, Shetty AK, Tshimanga M, Woelk G, Stranix-Chibanda L, Rusakaniko S. Group problem-solving therapy for postnatal depression among HIV-positive and HIV-negative mothers in Zimbabwe. J Int Assoc Provid AIDS Care. 2014;13(4):335–41.

65. Cooper PJ, Murray L, Wilson A, Romaniuk H. Controlled trial of the short-and long-term effect of psychological treatment of postpartum depression: 1. Impact on maternal mood. Br J Psychiatry. 2003;182(5):412–9.

66. Gregoire A, Kumar R, Everitt B, Studd J. Transdermal oestrogen for treatment of severe postnatal depression. Lancet. 1996;347(9006):930–3.

67. Kashani L, Eslatmanesh S, Saedi N, Niroomand N, Ebrahimi M, Hosseinian M, et al. Comparison of saffron versus fluoxetine in treatment of mild to moderate postpartum depression: a double-blind. Randomized clinical trial. Pharmacopsychiatry. 2017;50(2):64–8.

68. Logsdon MC, Wisner K, Hanusa BH, Phillips A. Role functioning and symptom remission in women with postpartum depression after antidepressant treatment. Arch Psychiatr Nurs. 2003;17(6):276–83.

69. Milgrom J, Gemmill AW, Ericksen J, Burrows G, Buist A, Reece J. Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: a randomised controlled trial. Aust N Z J Psychiatry. 2015;49(3):236–45.

70. Milgrom J, Holt CJ, Gemmill AW, Ericksen J, Leigh B, Buist A, et al. Treating postnatal depressive symptoms in primary care: a randomised controlled trial of GP management, with and without adjunctive counselling. BMC Psychiatry. 2011;11:95.

71. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. Br J Clin Psychol. 2005;44(Pt 4):529–42.

72. Misri S, Reebye P, Corral M, Mills L. The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: a randomized controlled trial. J Clin Psychiatry. 2004;65(9):1236–41.

73. Morrell CJ, Warner R, Slade P, Dixon S, Walters S, Paley G, et al. Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial. Health Technol Assess. 2009;13(30):1–153 (iii–iv, xi–xiii).

74. Mulcahy R, Reay RE, Wilkinson RB, Owen C. A randomised control trial for the effectiveness of group Interpersonal Psychotherapy for postnatal depression. Arch Womens Ment Health. 2010;13(2):125–39.

75. Ngai FW, Wong PW, Leung KY, Chau PH, Chung KF. The effect of telephone-based cognitive-behavioral therapy on postnatal depression: a randomized controlled trial. Psychother Psychosom. 2015;84(5):294–303.

76. O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. Arch Gen Psychiatry. 2000;57(11):1039–45.

77. Pinheiro RT, Botella L, de Avila Quevedo L, Pinheiro KAT, Jansen K, Osório CM, et al. Maintenance of the effects of cognitive behavioral and relational constructivist psychotherapies in the treatment of women with postpartum depression: a randomized clinical trial. J Constr Psychol. 2014;27(1):59–68.

78. Prendergast J, Austin M-P. Early childhood nurse-delivered cognitive behavioural counselling for post-natal depression. Australas Psychiatry. 2001;9(3):255–9.

79. Rojas G, Fritsch R, Solis J, Jadresic E, Castillo C, González M, et al. Treatment of postnatal depression in low-income mothers in primary-care clinics in Santiago, Chile: a randomised controlled trial. Lancet. 2007;370(9599):1629–37.

80. Sharp D, Chew-Graham C, Tylee A, Lewis G, Howard L, Anderson I, et al. A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the RESPOND trial. Health Technol Assess. 2010;14(43):1–153.

81. Wisner K. Estradiol for postpartum depression: translational challenges. Neuropsychopharmacology. 2014;39:S1–111.

82. Wisner KL, Perel JM, Peindl KS, Hanusa BH, Findling RL, Rapport D. Prevention of recurrent postpartum depression: a randomized clinical trial. J Clin Psychiatry. 2001;62(2):82–6.

83. Yonkers KA, Lin H, Howell HB, Heath AC, Cohen LS. Pharmacological treatment of postpartum women with new onset major depressive disorder: a randomized controlled trial with paroxetine. J Clin Psychiatry. 2008;69(4):659.

84. Zlotnick C. Effectiveness of sertraline alone and interpersonal psychotherapy alone in treating women with postpartum depression. ClinicalTrials.gov; 2016. https://clinicaltrials.gov/ct2/show/study/NCT00602355?sect=X4301256. Accessed 7 Nov 2018

85. Ahmadpanah M, Nazaribadie M, Aghaei E, Ghaleiha A, Bakhtiari A, Haghighi M, et al. Influence of adjuvant detached mindfulness and stress management training compared to pharmacologic treatment in primiparae with postpartum depression. Arch Womens Ment Health. 2018;21(1):65–73.

86. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2017. https://www.R-project.org/. Accessed 27 Mar 2018.

87. Rücker G, Schwarzer G, Krahn U, König J. Netmeta: Network meta-analysis using frequentist methods. 2018. https://CRAN.R-project.org/package=netmeta. Accessed 19 Jun 2018.

88. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.

89. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health. 2012;15(6):940–7.

90. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raon: Chapman & Hall/CRC; 1993.

91. Therneau TM, Crowson CS, Atkinson EJ. Adjusted survival curves. 2015. https://cran.r-project.org/web/packages/survival/vignettes/adjcurve.pdf. Accessed 7 Nov 2018.

92. Alexander JL. Quest for timely detection and treatment of women with depression. J Manag Care pharm. 2007;13(9 Suppl A):S3–11.

93. Flynn HA, Davis M, Marcus SM, Cunningham R, Blow FC. Rates of maternal depression in pediatric emergency department and relationship to child service utilization. Gen Hosp Psychiatry. 2004;26(4):316–22.

94. Kauppi A, Kumpulainen K, Vanamo T, Merikanto J, Karkola K. Maternal depression and filicide-case study of ten mothers. Arch Womens Ment Health. 2008;11(3):201–6.