1. Introduction

1.1. Rationale

Sleep deprivation alone is associated with one of the most rapid treatment responses in psychiatry (Wirz-Justice & Van den Hoofdakker, 1999). It may improve mood in depressed patients in a single 24-hour period, with responders usually achieving euthymia by the end of the intervention. Boland et al. (2017) conducted a meta-analysis on the antidepressant effects of sleep deprivation alone and found that about 50% of depressed patients respond rapidly within 24 hours. They did not find significant differences in clinical effectiveness based on the type of sleep deprivation used, demographics of patients, medication status, or whether the depression was unipolar or bipolar. The main problem of acute sleep deprivation is the high relapse rates following a subsequent night of sleep or even a simply diminished effect by taking naps during sleep deprivation. As such, researchers in the 1990’s started to experiment with additional components aimed at sustaining clinical effectiveness following sleep deprivation. Common additions include repeating the cycle of sleep deprivation, adding sleep phase advance and augmentation with bright light therapy in the morning. When combined with sleep deprivation, these three steps constitute a procedure called ‘Triple Chronotherapy’ which is a type of chronotherapy aimed at resetting and stabilising the patients’ circadian rhythms. Depression is associated with circadian rhythm and sleep disturbances, and the hypothesis is that chronotherapy may, at least in part, resynchronise the circadian rhythms and stabilise sleep timing (Wirz-Justice & Benedetti, 2019). Multi-system mechanisms may underlie these treatments, acting to regulate neurotransmitters such as monoamines (e.g. dopamine, serotonin) or hormones such as melatonin so that they are released at the right time of day. It may act on the regulation of cortical neuroplasticity and glial cell cycles; on the peripheral hormonal rhythms; and directly, on brain areas which contribute to the control of affect, and the reactivity to external stimuli of the HPA axis (Wirz-Justice & Benedetti, 2019).

There are several variations of sleep deprivation and combinations of bright light therapy reported in the literature. It is worth noting that not all variations or
combinations may fit the definition of Triple Chronotherapy outlined above, and in this meta-analysis we regard Triple Chronotherapy as within the wider category of chronotherapy. These include total sleep deprivation (e.g. deprived of sleep for the entirety of 36 hours) or partial sleep deprivation (e.g. being allowed to sleep in the first half of the night followed by a prolonged period of wakefulness). Sleep deprivation may be prescribed for a single night followed by advancing the phase of sleep over 4 days. Blue blocking (amber) glasses may also be used to encourage release of melatonin in the four hours prior to sleeping in the four days of sleep phase advance after one night of sleep deprivation. Alternatively sleep deprivation occurs 3 times over 6 days in which each cycle of sleep deprivation is followed by a night of recovery sleep. Whichever method of sleep deprivation is used, the procedure is usually combined with bright white light in the morning for sustained therapeutic effects.

1.2. Objectives
This meta-analysis aimed to further investigate the antidepressant effects of chronotherapy, which in this case are defined as any combinations of sleep deprivation followed by sleep phase manipulation and/or additional bright light therapy. We focused on studies that use sleep deprivation as the primary intervention. The questions addressed were:
1) How effective is chronotherapy in the treatment of a depressive episode in unipolar and bipolar disorder compared to other routine treatments?
2) How sustainable are the treatment effects of chronotherapy at follow-up, as indexed by a reduction in depressive symptoms?

In this meta-analysis we included both randomised controlled trials (RCTs) and open label case series. This is partly because of the limited number of studies available with this novel treatment. Although conclusions from case series are problematic due to the lack of a control group, publication bias and because they may represent a maximum of results that can be obtained, they can potentially support the findings from RCT's and provide additional information as the design is more flexible (Chambers, Rodgers and Woolacot, 2009) Data from RCTs and case series were analysed separately, including within-group change for the case series and both between-group and within-group change for the RCTs to allow comparison between both kinds of study. .
2. Methods

2.1. Protocol and registration
The review protocol for the current meta-analysis had been pre-registered on PROSPERO (Protocol registration number: CRD42019127228).

2.2. Eligibility criteria
We adhered to the following eligibility criteria for inclusion of studies 1) studies must be in the English language, published in peer-reviewed journals and have had full-text versions at the time of search (i.e. not abstracts only); 2) studies must either be case series or randomised controlled trials, whereas reviews and single-case reports were excluded; 3) studies must include experimentally induced partial or total sleep deprivation followed by sleep phase advance or delay with or without light therapy in adult populations; 4) patients must be diagnosed with a depressive episode (unipolar/bipolar) at the time of the study; and 5) observer-rated or self-rated outcome measures must have been used, and studies only using visual analogue scales were excluded.

2.3. Information sources
A comprehensive search of the literature was carried out on 8th August 2018 using PubMed, EMBASE, PsychINFO and CINAHL databases with additional sources such as reference lists of existing studies and Google Scholar records. No date limits were imposed in the search as the earliest record was found in the year 1985.

2.4. Search
Key words used were a combination of ‘Chronotherapy’ and ‘Depression’, which resulted in 88 records being retrieved (PRISMA flowchart). An additional seven studies were added from other sources, such as reference lists of existing studies. ‘Grey literature’ (that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers) was also searched via the ‘Open Grey’ website.
2.5. Study selection

After screening (excluding reviews and non-experimental studies) 25 records remained and after applying the eligibility criteria outlined above nine studies were excluded, one due to being single-case report (Gottlieb & Terman, 2012); one due to study population being minors only (Gest et al., 2016); one due to having only a visual analogue scale as an outcome measure (Colombo et al., 2000); one due to having qualitative methods only (Kragh, Moller, et al., 2017); two due to being crossover studies and having no follow-up data at one week (van den Burg, Bouhuys, van den Hoofdakker, & Beersma, 1990; Wehr, Rosenthal, Sack, & Gillin, 1985), whereas the remainder were repeat or overlapping data of a study reported in separate publications (Kragh et al., 2018; Martiny et al., 2013; Martiny et al., 2015). As such, a total of 16 studies were included in the final analysis. Of these 16, four were randomised controlled trials (Kragh, Martiny, et al., 2017; Martiny et al., 2012; Neumeister et al., 1996; Wu et al., 2009) and 12 were open labelled case series (Benedetti et al., 2001; Benedetti et al., 2005; Benedetti et al., 2014; Berger et al., 1997; Dallaspezia, Suzuki, Clara, Colombo, & Benedetti, 2018; Dallaspezia & van Jaarsveld, 2016; Echizenya, Suda, Takeshima, Inomata, & Shimizu, 2013; Moscovici & Kotler, 2009; Riemann et al., 1999; Sahlem et al., 2014; Voderholzer et al., 2003; Vollmann & Berger, 1993).

2.6. Data collection process/Data items

Each of the 16 articles was reviewed and the following data were extracted: publication year, country, patient type and sample size (unipolar, bipolar or mixed, intervention/control for RCTs), medication status and number of patients receiving each type of medication (e.g. antidepressants, antipsychotics), type and duration of sleep deprivation (e.g. partial or total), control groups for RCTs, additional light therapy, primary outcome measure, response criterion, percentage of responders (intervention/control for RCTs), follow-up periods, adverse effects from sleep deprivation and adverse effects from light therapy. Two authors (Benedetti et al., 2001;
Benedetti et al., 2005; Benedetti et al., 2014; Wu et al., 2009) were contacted for additional data. Data were extracted and checked by two of the authors (C.H. and D.V.).

2.7. Risk of bias in individual studies
We assessed the quality and risk of bias of the four RCTs according to the criteria outlined in the assessment tool developed by the Cochrane Collaboration (Higgins et al., 2011). These included random sequence generation, allocation concealment, blinding of participants and personnel (single-blind in all cases), blinding of outcome assessment, incomplete outcome data, and selective reporting. Three out of four RCTs included in this meta-analysis were deemed as ‘low risk’ according to these criteria, with only one showing ‘some concerns’ (Neumeister et al., 1996).

2.8. Summary measures
For both RCT’s and case series we used mean change from baseline to calculate ES, using Hedge’s g to estimate standardised mean differences. Hedge’s g corrects Cohen’s d for small sample bias. Cohen’s d in this case is calculated by the mean difference in change divided by the standard deviation of change scores within each study. The SD of mean change is dependent on the standard deviation of pre- and post- scores and the correlation coefficient between them. However, for this set of studies, the correlation between baseline and follow-up was unknown, and therefore we used a coefficient of 0.5 as a reasonable default. Effects of treatment for RCTs were the difference in change scores between the treatment and control groups. For the case series, as there was no control condition, treatment effects were represented by the change in outcome scores from baseline. Effect sizes were calculated at different timepoints, namely 1-2 days, 5-7 days and 7-9 weeks, with 5 – 7 days chosen as the main timepoint of interest. When calculating effect sizes, only one outcome measure [various versions of the Hamilton Depression Rating Scale (HAM-D)] or the Beck Depression Inventory (BDI) was used per study. To aid interpretation, we also calculated responder rates in 2 ways. Firstly by calculating the pooled odds ratio for responding across treatment condition for the RCTs, and secondly by estimating the overall probability of response within treatment groups for the RCTs and case series. Note that we present these as descriptive statistics (pooled point estimates and 95% CIs) to aid interpretation.
2.9. **Synthesis of results**

A random effects model was used as we expected heterogeneity of effect sizes. Pooled effect sizes were calculated using weights derived according to the inverse-variance method.

To describe the heterogeneity of the effect sizes we used the $I^2$ statistic which is derived from the Q statistic for heterogeneity. This ranges from 0% to 100%, with 0% to 25% indicating little observed heterogeneity, 50% moderate, and towards 75% and above high to very high. The 95% CI for $I^2$ was calculated using the heterogi module in STATA. Statistical inference for the pooled treatment effect followed from calculating Z-statistics and associated $p$ value. Review Manager 5.3 was used to pool estimates for standardised mean differences whilst the package metafor in R 2.5 was used to calculate rates of response and odds ratios.

2.10. **Risk of bias across studies**

We used RoB 2 (the latest Cochrane risk of bias tool) to ascertain risk of bias across RCTs, and found most studies to be of low risk (Supplementary Materials). Publication bias (such as those indexed by a funnel plot) is difficult to assess given the low number of RCTs included; however, a funnel plot had been produced for the case series and can be found in Supplementary Materials.

3. **Results**

3.1. **Study selection**

Four RCTs and 12 open-label case series were considered to fit the inclusion criteria (PRISMA flow diagram, Figure 1).

3.2. **Study characteristics**

Tables 1 and 2 show the characteristics of included RCTs and case series, respectively. Almost all studies (except one (Dallaspezia & van Jaarsveld, 2016)) were
carried out in a psychiatric in-patient setting. The majority of studies defined treatment response as a reduction of 50% or more in the primary outcome measure for depressive symptoms. For RCTs, three out of four were conducted in Europe and one in the United States whereas for case series, ten out of twelve were carried out in Europe with one in Asia and one in the United States. No chronotherapy study has been conducted in the United Kingdom. In terms of patient population, two out of 4 RCTs had a mixture of unipolar and bipolar patients whereas three out of 12 case series had mixed patients. In total there were 150 unipolar and 58 bipolar patients for RCTs, 174 unipolar and 330 bipolar patients for the case series. The numbers of patients included in each case series varied considerably, from a minimum of 10 to a maximum of 141. All patients enrolled in RCTs were on medication at the time of the treatment and 96 patients from only three case series had no medication.

The most prominent difference between RCTs was the length of maximum follow-up, ranging between seven days to nine weeks. Secondly, the duration and type (full/partial) of sleep deprivation, whether it is repeated over 3 cycles or sleep phase advance varied considerably. There are significant differences between the addition of medication, and the use or duration of light therapy that was added to stabilise the effect of total sleep deprivation on the first night. Thus the first case series (1993-2003) did not include any additional light therapy. Two RCTs and three case series used repeated cycles of sleep deprivation. Although all four RCTs used light therapy, the intensity ranged from 3,000 to 10,000 lux with varying durations of exposure.

Adverse effects were not reported systematically by all RCTs and when these were included, percentages of patients affected seemed to vary greatly between studies with one more recent studies reporting more frequent side effects (e.g. (Kragh, Martiny, et al., 2017)). Some RCTs also used modified versions of outcome measures suited to the study, such as a 24-item HAM-D which excluded sleep related items (21-item HDRS-NOW). Control treatments in the four RCTs were entirely different from one another and there were no measures of credibility in the control treatment. A similar pattern was observed for case series as well, with the follow-up period as the most varied parameter between studies which ranged between five days and three months. One case series used a self-report measure, the BDI (Dallaspezia & van Jaarsveld, 2016) whereas all others used observed rated measures (different versions of the HAM-D). Lastly, there has been a trend in case series to add bright light therapy with
reported intensity now up to 10,000 lux of bright white light at 30cm distance for 30 minutes. However, no study reported light intensity and light spectrum at the level of the eye, nor light orientation (downwards, upwards, or lateral directed).

Insert Tables 1 and 2 about here

3.3. *Effects of chronotherapy on depressive symptoms (Day 5 – 7)*
Effect sizes and 95% confidence intervals for RCTs are shown in Table 3 and Figure 2. All four RCTs showed an effect favouring experimental intervention (chronotherapy) compared to control treatments at Day 5 - 7, with a pooled effect size of $g = 0.62$, 95% CI 0.23 – 1.01. Heterogeneity was moderate ($I^2 = 40\%$, 95% CI 0%, 79%) although the 95% CI does not rule out low or high values. Figure 5 clearly shows effects favouring Day 5 - 7 compared to baseline measurements in case series ($g = 1.79$, 95% CI 1.50 – 2.08). Heterogeneity however was high ($I^2 = 68\%$, 95% CI: 42%, 83%).

Insert Table 3 about here

Insert Figure 2 about here

3.4. *Immediate effects of chronotherapy on depressive symptoms (Day 1 – 2)*
Table 4 displays the effect sizes and 95% confidence intervals for separate intervention and control groups for RCTs immediately post-treatment (Day 1 or 2) and at various follow-up points. It is demonstrated that the strongest effects were apparent after Day 1 or 2 compared with baseline ($g = 2.23$, 95% CI 0.71 – 3.75). Table 5 shows a similar pattern with open-label case series at Day 1 or 2 ($g = 1.75$, 95% CI 1.46 – 2.04).

3.5. *Effects of chronotherapy on depressive symptoms at longer-term follow-up points (Week 7 – 9)*
Tables 3-4 demonstrate that beneficial effects tended to attenuate at later follow-up points for RCTs which potentially suggest a somewhat diminishing therapeutic effect of chronotherapy over time ($g = 0.35$, 95% CI -0.08 – 0.77). However, at Week 7 - 9, the effects still strongly favoured post-test follow-up in the experimental intervention
group (Figure 3) compared with baseline measurements in the same group, whereas the effect sizes for the control group (Figure 4) were much smaller (although still favouring follow-up). Heterogeneity was low ($I^2 = 15\%$) amongst the experimental group whereas in the control group it was considered to have no heterogeneity ($I^2 = 0\%$). However, it does not show a steady diminishment of effects at follow-up points for open-label case series. Three out of the 12 studies conducted a follow up between 4 and 12 weeks and found that the effects were generally sustained (ranging from $g = 1.53$, 95% CI 0.96 - 2.10 to $g = 2.84$, 95% CI 1.76 – 3.92).

Insert Figures 3 and 4 about here

Insert Table 4 about here

Insert Table 5 about here

Insert Figure 5 about here

3.6. Effects of chronotherapy in unipolar versus bipolar patients (Day 5 – 7)
As a secondary analysis we were interested in whether chronotherapy was more effective in unipolar or in bipolar depression patients. After excluding mixed samples we calculated within-group effect sizes for both RCTs (chronotherapy group) and case series which included unipolar (N = 116, 6 studies) only or bipolar (N = 342, 5 studies) only at Day 5 – 7, and found similar effect sizes between the effectiveness of chronotherapy in unipolar ($g = 2.16$, 95% CI 1.57 – 2.75) or bipolar ($g = 1.87$, 95% CI 1.46 – 2.28) patients.

3.7. Response rates to chronotherapy
It should be noted that Neumeister et al. (1996) included only responders to sleep deprivation in the study population and was excluded from the responder analysis because of missing data in the control group. They also had an exceptionally high (100%) response rate in the intervention group as they only included responders. Therefore, we decided not to include this study when calculating weighted means and yielded 33.0% (95% CI 12.5% - 63.0%) as the response rate for the chronotherapy
group in RCTs and 1.5% (95% CI 0% - 52%) for patients in the control group in RCTs (OR = 7.58, 95% CI 2.03 - 28.28). Open labelled case series on the other hand had a weighted mean response rate of 61.6% (95% CI 54.4% - 68.3%).

4. Discussion

4.1. Summary of evidence

Pooled estimates from both RCT’s and case series show that chronotherapy appears to be effective, providing feasible and acceptable interventions for both unipolar and bipolar depression. Moreover, therapeutic response is usually manifest within a week. However, the severity and chronicity of depressive symptoms are thought to play a major role in treatment response, which may be attenuated in more treatment refractory cases. For RCTs, the standardised effect sizes (Hedge’s $g$) are large to very large favouring the chronotherapy except Kragh et al. (Kragh, Martiny, et al., 2017) which has a small to moderate effect size in favour of the chronotherapy. One curious finding is that even for control groups the effect sizes are moderate to large. Potential reasons for these are described in the limitations below, including a possible placebo effect. For case series, again we have found effect sizes at least in the moderate range when evaluating changes within-group longitudinally. However, the follow-up period for these case series is usually short and as such it is difficult to assess the sustainability of chronotherapy.

One study focused on reduction of suicide risk in inpatients (27). None of the studies reported any significant side effects during chronotherapy except for Kragh et al (16). This may represent under-reporting. Only 4/ 388 (1%) of bipolar patients switched to mania. The large majority of bipolar patients were taking lithium, which may contribute to preventing a switch.

4.2. Limitations and recommendations

The number of RCTs included in this meta-analysis is small and all have small sample sizes meaning inflated effect sizes can be a concern. Case studies lack a control group and thus it is not possible to separate out intervention and placebo effects. There is thus the need for larger, high quality trials to establish a stronger evidence-base.
Specific limitations to all of the studies included (RCTs and case series) is that (a) they nearly all included in-patients and the results may not be generalisable to out-patients, (b) nearly all the subjects lacked credibility and expectation ratings before receiving treatment (e.g.(Rief et al., 2009), in other words it remained unknown as to whether the patients expected the intervention given to be effective. This kind of placebo effect may partially explain why even for control conditions their effect sizes are still large (c) they all had different active controls (all were on medication and one with exercise, without standardisation of procedures across studies), (d) the short follow-up periods for most of the studies exclude the potential of further investigating the durability or any long-term benefits of chronotherapy. In addition, patients under 18 were not included in the meta-analysis. However, despite being very few in number, two studies in adolescents have shown moderate to strong effects of chronotherapy in young people with major depression, including one RCT (Gest et al., 2016) and a recent case series (Hurd, Herrera, Brant, Coombs, & Arzubi, 2019).

Another limitation is the heterogeneity of patient populations with depression compounded with a lack of clear definitions for response/remission across studies. Indeed, the \( I^2 \)-squared statistics across all studies indicate moderate to high heterogeneity. There is also heterogeneity in the chronotherapy intervention to stabilise the effect of total sleep deprivation. We do not know which responders to total sleep responders can be best stabilised by repeating the cycle of Total Sleep Deprivation, by adding Sleep Phase Advance or by adding Bright Light Therapy or a combination of the above.

The aforementioned study by Kragh et al. (2017) had a significantly higher proportion of patients on antipsychotics and antidepressant polypharmacy. Patients who had an evening chronotype (i.e. people who go to bed and get up later), a diurnal variation in mood (i.e. those who regularly feel worse in the morning), and higher levels of observed activity are reported to be predictors for response to chronotherapy (Bouhuys, Beersma, & Van den Hoofdakker, 1988; Kragh et al., 2018; Reinink, Bouhuys, Wirz-Justice, & Van den Hoofdakker, 1990). Conversely, the length of treatment resistance to other therapies (including antidepressants and mood stabilisers) appears to be also detrimental to the response to chronotherapy (Beneditti et al., 2005). The case series had a higher proportion of bipolar patients and higher effect sizes within the group. This may be partly because bipolar patients might have a better response than unipolar disorder, or because they do not have a control group.
The standard mean differences in depression scores at 7 – 9 weeks suggest there may be improvement with chronotherapy, but as the confidence intervals overlap with each other, no firm conclusion about the sustainability of therapeutic effects can be drawn. This may be because studies are underpowered and a larger study with longer follow up would be worthy of further research which requires larger studies, preferably in out-patients, where the large majority of depression is treated. Follow up is required after at least six months with both blind observer-rated and subjective-rated measures on depression and quality of life. A credible control is needed with ratings in both groups. The bright light therapy (for a meta-analysis, see (Golden et al., 2005)) should probably be 10,000 Lux at 30cm for 30 minutes with the timing optimised by the Morningness-Eveningness Questionnaire. This is a similar recommendation for treating seasonal affective disorder and non-seasonal depression (Lam et al., 2016).

The observation that there was no major difference in the effect sizes between unipolar and bipolar patients meant that both groups could be offered chronotherapy, but it is probably still important to have separate groups for RCTs if only to compare effect size to other treatments. Cost-effectiveness studies are also required as this has not been investigated. Further studies on acceptability and choice over alternatives are required.

5. Conclusion
To the best of our knowledge, this is the first meta-analysis concerning the use of combinations of chronotherapy. Current evidence suggests that chronotherapy is effective and generally well-tolerated treatments for major depressive episodes in both unipolar depression and bipolar disorder. Compared with routine treatments such as medication, talking therapy or exercise, chronotherapy has the added benefit of rapid treatment response in addition to a highly favourable side effect profile. Further, there is no significant switching to hypomania in bipolar disorder following chronotherapy. However, there is significant heterogeneity across studies especially with open-labelled case series, which clearly indicates the necessity to standardise the use of chronotherapies in major depression. Further research is needed to identify circadian rhythm markers of response and more high-quality studies, especially RCTs employing specific types of chronotherapy (e.g. triple chronotherapy) tailored to each individual’s chronotype with longer follow-up periods or those conducted in a community setting, are therefore urgently needed.
References


