

Title: Sham sleep feedback delivered via actigraphy biases daytime symptom reports in people with insomnia: implications for insomnia disorder and wearable devices

Short-title: Sham feedback biases insomnia-related symptom reports

Authors: Gavriloff, D.,^{a,b} (DClinPsych), Sheaves, B.,^{a,b,c} (DClinPsy), Juss, A.,^a (BSc), Espie, C.A.,^a (PhD), Miller, C.B.,^d (PhD), Kyle, S.D.^a (PhD)*

Affiliations:

^a Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford

^b Oxford Health NHS Foundation Trust

^c Department of Psychiatry, University of Oxford

^d CIRUS, Centre for Sleep and Chronobiology, Woolcock Institute of Medical Research, The University of Sydney, Australia

*Correspondence should be addressed to:

Simon D. Kyle, PhD
Sleep and Circadian Neuroscience Institute (SCNi)
Nuffield Department of Clinical Neurosciences
Sir William Dunn School of Pathology
University of Oxford
South Parks Road
Oxford
OX1 3RE

Tel: 01865 618675

Email: simon.kyle@ndcn.ox.ac.uk

Disclosure statement: Financial support was provided by Oxford Health NHS Foundation Trust, a Wellcome Trust strategic award to the Sleep and Circadian Neuroscience Institute and the Dr Mortimer and Theresa Sackler Foundation.

Total words: 5944 (Summary, Introduction, Methods, Results, Discussion, References)

Total references: 39

Author contributions:

DG, SDK and BS designed the study, analysed and interpreted the data and wrote the research paper. DG and AJ ran the experimental protocols. CBM supported statistical analysis and critically reviewed the paper. CE critically reviewed the paper and supported interpretation of findings.

Summary

This study investigated whether providing sham feedback about sleep to individuals with insomnia influenced daytime symptom reports, sleep-related attentional bias and psychomotor vigilance. Sixty-three participants meeting DSM-5 criteria for insomnia disorder were recruited from the community. Following baseline assessments and actigraphy briefing, participants were randomised to receive next-day sham feedback on sleep quality ('positive' versus 'negative' sleep efficiency condition). Feedback was delivered at habitual rise-time using an integrated actigraphy-diary watch to simulate wearable device behaviour. Participants completed symptom reports immediately before receiving feedback, and at 12:00 and 15:00 hrs, using the experience sampling method. Following this they returned to the lab in the evening to complete symptom reports and computerised tests of sleep-related attentional bias and basic psychomotor vigilance. Participants randomised to negative feedback ($n = 32$) evidenced impaired daytime function (decreased alert cognition [$d = .79$], increased sleepiness/fatigue [$d = .55$]) in the evening compared with those given positive feedback ($n = 31$). Within-day trajectories revealed that the positive feedback group, relative to the negative feedback group, displayed a significantly greater increase in positive mood and alert cognition (from rise-time to 12:00 noon), and significantly greater decrease in sleepiness/fatigue. There were no significant between-group differences on measures of sleep-related attentional bias [$d = .20$] or psychomotor vigilance [$d = .12$]. This controlled experiment shows that sham feedback about sleep biases appraisal of daytime symptoms, highlighting a pathway connecting sleep misperception with daytime features of insomnia. Findings have important implications for wearable devices that claim to measure "objective" sleep yet may provide inaccurate data relative to gold-standard measurement.

Keywords: misperception; functioning; attention bias

Introduction

A reliable yet perplexing finding in insomnia disorder is the tendency for patients to overestimate sleep onset latency (SOL) and underestimate total sleep time (TST) relative to objective measures (Chambers & Keller, 1993; Harvey & Tang, 2012). Discrepancy in daytime impairment is also common, whereby reports of impaired function are typically more pronounced than objective tests of performance (Fortier-Brochu *et al.*, 2012). While the precise role of misperception in the development and maintenance of insomnia is yet to be elucidated, Harvey (2002) places particular emphases on key cognitive processes. Within a feedback model, it is argued that misperception provides an activating process for psychophysiological arousal, distress and safety-seeking behaviours. These are theorised to drive selective attention for sleep-related threat and confirm perceptions of deficit, serving to further exacerbate negative cognitions and engender *actual* deficits in sleep and daytime function (Harvey, 2002).

Semler and Harvey (2005) used sham feedback to examine the relationship between sleep (mis)perception and daytime functioning in people with insomnia. For three mornings participants were given feedback on 'sleep quality', which was apparently computed from a wrist-worn actigraph and delivered through a bedside display device. In fact, they received sham feedback that was pre-programmed into the display device. Negative feedback was associated with greater endorsement of negative sleep-related thoughts, monitoring for sleep-related threat, use of safety behaviours, and enhanced sleepiness relative to days when participants received positive feedback.

In the current era of wearable technology the impact of digitally delivered feedback on behaviour, cognition and health is of clinical and scientific importance (Patel *et al.*, 2015; Piwek *et al.*, 2016). In relation to sleep, there is mounting concern that so-called "sleep tracking" devices have limited evidence of validity or reliability (Behar *et al.*, 2013; Ko *et al.*, 2015; Lee & Finkelstein, 2014; Van den Bulck, 2015). For example, one recent study found that a leading commercial device markedly underestimated total sleep time and sleep efficiency relative to polysomnography in patients with

major depression (Cook *et al.*, 2017). If these devices provide inaccurate feedback about sleep – particularly if overestimating the magnitude of disturbance – they may heighten or reinforce sleep misperception and thereby trigger, maintain or exacerbate sleep disruption and daytime symptomatology.

The current study sought to test the hypothesis that giving participants with insomnia sham ‘objective’ feedback about their sleep would bias daytime functioning and sleep-related attentional processes. Critically, we devised an experimental protocol that could plausibly simulate the everyday use of wearable devices in the real-world. It was hypothesised that, relative to positive feedback, participants given negative feedback would report impaired daytime function, increased vigilance for sleep stimuli (sleep-related attentional bias) and lower levels of psychomotor vigilance.

Methods

Participants

Participants with insomnia were recruited from the community using social media, email and poster advertisements. For study inclusion participants were required to be ≥ 18 years of age and report clinically significant insomnia. Exclusion criteria were: indication of clinically significant depression or anxiety; use of a sleep tracking device more than once in the past month; a psychiatric diagnosis; use of prescribed medication for sleep; or typical sleep patterns that lay outside the hours of 22:00 to 9:00. There was no exclusion based on disorders of physical health.

A score of ≤ 16 on the Sleep Condition Indicator (SCI; Espie *et al.*, 2014) was used to indicate clinically significant insomnia. The SCI is modelled on DSM-5 criteria for insomnia and has been shown to possess excellent internal consistency ($\alpha = 0.86$) and convergent validity with the Pittsburgh Sleep Quality Index (Buysse *et al.*, 1989) and Insomnia Severity Index (Bastien *et al.*, 2001). Participants were required to score < 11 on the Patient Health Questionnaire-9 (PHQ-9; Kroenke *et al.*, 2001; Manea *et al.*, 2012) and < 10 on the Generalised Anxiety Disorder-7 (GAD-7; Spitzer *et al.*, 2006) to rule out

probable depression or an anxiety disorder. The Brief Screen for Sleep Disorders (BSSD; Wilson *et al.*, 2010) and Berlin Questionnaire (Netzer *et al.*, 1999) were used to screen for sleep disorders other than insomnia, including likely narcolepsy, sleep breathing disorder, restless legs syndrome, circadian rhythm disorder or parasomnia.* The final sample comprised 63 participants (49 females, 14 males; mean age = 44.4, standard deviation [SD] = 14.5 years; Figure 1).

(Insert Figure 1 about here)

Measures

Sleep

Subjective sleep was assessed using the Core Consensus Sleep Diary (CCSD; Carney *et al.*, 2012). The diary permitted quantification of subjective total sleep time and sleep efficiency prior to the delivery of “objective” feedback.

Participants were given an integrated actigraph watch (PRO-Diary Actiwatch; CamNtech Ltd.), a small motion-sensitive device that contains an accelerometer and is worn around the wrist. This model of actiwatch allows for time-stamped data entry through a touchscreen display and has been used in previous research to examine inter-relations between sleep disturbance and daytime symptoms in clinical samples (Mulligan *et al.*, 2016). Participants entered sleep diary and symptom data (DISS; Buysse *et al.*, 2007) directly into the actiwatch on the second day. The watch also served as the delivery mechanism for the sham “objective” feedback (see *procedure* section).

Daytime function

The Daytime Insomnia Symptom Scale (DISS; Buysse *et al.*, 2007) is a 20-item self-report measure that was used to assess subjective daytime functioning. Designed for ecological momentary assessment, items are distributed across four validated factors: *alert cognition* (forgetful [reverse scored], clear-

* Initially the BSSD was used to screen for sleep disorders other than insomnia but this was replaced by the Berlin Questionnaire (focusing on apnoea exclusion only) at one month into the ten-month recruitment process.

headed, able to concentrate, how much of an effort is it to do anything, alert), *negative mood*, (anxious, stressed, tense, sad, and irritable), *positive mood* (relaxed, energetic, calm, happy, and efficient), and *sleepiness/fatigue* (fatigued, sleepy, and exhausted). Each item is measured on a visual analogue scale (VAS) ranging from “very little/poorly/bad” to “very much/good/well”. Two versions of the measure were used: a paper version with a 10cm VAS, and a digital version for use on the actiwatch using a scrollable 10cm VAS. The DISS has been used in several other studies to profile sleep-related daytime dysfunction (Buysse *et al.*, 2007; Miller *et al.*, 2013; Russell *et al.*, 2016).

Computerised task performance

The computerised tasks were administered using a Dell Latitude D351 laptop with a 39cm display (diagonal) and accompanying response box (Model RB-400; Cedrus Corporation; San Pedro, CA).

Dot-probe task

The dot-probe task was used to measure sleep-related attentional bias, having previously been found to be sensitive to insomnia (MacMahon *et al.*, 2006; Harris *et al.*, 2015). Participants completed four practice trials followed by 160 experimental trials. Pairs of words (neutral and sleep-related words matched for length and frequency of common usage) are presented in a randomised order (see supplement for list of words used). Each trial consists of a fixation cross appearing at the centre of the screen for 500 milliseconds. This is then followed by a pair of words, one appearing above and one appearing below the position of the cross, which are visible for 500 milliseconds. After this, a dot-probe in the form of an asterisk, appears in either the upper or lower position and remains in view until the participant presses the correct response (either upper or lower button on a response box). Fifty percent of trials contain a sleep-related word with an equal probability of the probe replacing either the sleep-related or neutral word. Relative positions of each pair of words are counterbalanced within the experiment. Each trial is followed by an inter-trial interval of 1000 milliseconds. A mean interference index is calculated for each administration of the measure using the formula: [(Sleep Up-Dot Down mean RT + Sleep Down-Dot Up mean RT) – (Sleep Up-Dot Up mean RT + Sleep Down-Dot

Down mean RT)] / 2, with positive interference indices indicating heightened vigilance for sleep related words relative to neutral words, and negative indices indicating avoidance.

Psychomotor vigilance task

The psychomotor vigilance task (PVT) is a computerised measure of vigilant attention (Altena *et al.*, 2008; Dinges & Powell 1985). Participants complete five practice trials followed by 110 experimental trials. Each trial is comprised of an asterisk that appears in the centre of the screen. Participants are then told to 'left click' on a mouse as quickly as possible once they see the asterisk. Interval onset between asterisks varies from one to ten seconds. A PVT mean reaction time (RT) was calculated for each participant at each administration of the task, with higher scores indicative of slowed reactions.

Supplementary questions post-experiment

On completion of all experimental procedures and post-collection of outcome variables, participants were asked to rate two supplementary questions on feedback congruence and preoccupation. The former asked, "to what extent did you believe the feedback from the actiwatch reflected how you slept last night?" and the latter "to what extent did you find yourself thinking about your sleep during the day?". Participants were asked to rate both questions on a five-point scale ("not at all", "a little", "somewhat", "much", "very much"). Finally, participants were asked to answer ("yes" or "no") to: "last night did you use a sleep tracking device in addition to the actiwatch?"

Procedure

Participants were directed to an online survey platform where they were presented with the participant information sheet, gave electronic consent, and underwent questionnaire screening. Those who were deemed eligible were then contacted by the research team via telephone or email to arrange two brief visits to the laboratory on consecutive days (between 17:30 and 19:00). Consecutive visits to the lab were matched for time, to within one hour.

First laboratory visit (pre-manipulation)

Participants were asked to complete paper copies of the study consent form and three baseline measures (Figure 2). These comprised the DISS (subscales: *alert cognition*, *negative mood*, *positive mood*, *sleepiness/fatigue*), dot-probe task and PVT. After completing all measures, and in the presence of the researcher, participants were asked to read through two pages of brief psychoeducation about sleep and sleep efficiency. This briefing made clear that actigraphy is a reliable measure of sleep and wakefulness, and that sleep efficiency is a helpful way of conceptualising how well someone has slept. It specified that normal sleep efficiency is approximately 80-85% and anything above 90% is considered very good, while less than 75% is considered poor sleep efficiency, which has been linked to functional impairment. Participants were offered the chance to ask questions on what they had read and were then instructed in how to use the actiwatch and asked to wear it for 24 hours. A demonstration device was used to show participants how the watch operated. Participants were told that the actiwatch would deliver objective feedback on their sleep in the form of a sleep efficiency calculation at their habitual rise-time the next morning. In fact, the actiwatches were pre-programmed to deliver either “positive” or “negative” feedback (see below) depending on the experimental condition to which the participant was randomly allocated.

(Insert Figure 2 about here)

Participants were randomly allocated to receive either “positive” ($n = 33$) or “negative” ($n = 33$) feedback on their sleep by a member of the research team not involved with administration of experimental measures. Randomisation was achieved using a computerised random number generator and a block randomisation procedure with variable block size (2-6) to minimise prediction of future allocations. Group allocation was kept blind to the researcher administering the experimental tasks (DG) until study participation had ceased. The actiwatch was intended to reinforce the verisimilitude of the sleep feedback and to provide a delivery mechanism by which to do this, akin to other commercially available devices.

Day two

The initial actiwatch alarm was programmed to sound at the participants' pre-specified habitual rise-time. Participants completed the DISS and sleep diary via touch screen display before accessing the feedback from the actiwatch ("Your sleep efficiency was 91.4%/61.4%; this is in the range of VERY GOOD/VERY POOR sleep quality"), ensuring that daytime function and sleep appraisal could be captured pre-manipulation. Participants were then prompted by the actiwatch to complete the DISS again at 12:00 and 15:00 before returning to the laboratory in the evening.

Second laboratory visit (post-manipulation)

On arrival at the laboratory participants returned the actiwatch and again completed the DISS, dot-probe task and PVT. At this point they were also asked questions about feedback congruence (between device and sleep perception), feedback preoccupation and whether they had used any other sleep monitoring devices during the previous night. Participants were then comprehensively debriefed as to the real nature of the study and offered access to Sleepio™, a digital CBT programme for insomnia (Espie *et al.*, 2012).

Statistical analysis

An *a priori* sample size calculation (using G*Power 3 software; Faul *et al.*, 2007) was performed using an aggregated effect size of $d = .75$ taken from two experimental studies on attention bias (MacMahon *et al.*, 2006; Richardson *et al.*, 2014). With $\alpha = .05$ and power set at 80%, suggested sample size was $n = 29$ in each group. A total sample size of 64 would therefore allow for reasonable attrition of 10% ($n = 6$). Primary outcomes (daytime function [DISS], attention bias [dot-probe task], psychomotor vigilance [PVT]) were initially examined for between-group differences on evening two, using analysis of covariance (ANCOVA), covarying for baseline (evening one) scores. For within-day analyses, Linear mixed effect models were conducted on each DISS subscale to appraise differential group change from rise-time. Post-feedback scores (12:00 noon and 15:00 hrs) were modelled against randomised group, time-point, the interaction between randomised group and time point, and baseline (rise-time, pre-feedback) score. A random intercept for each participant was included in the model as a random effect

to account for repeated measures on the same participant. Linear mixed effect models were utilised to account for missing observations (Olsen *et al.*, 2012), which constituted 13 missed data-points out of a possible 189 within-day data points. Data were analysed using the SPSS Statistical Package Version 22 (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) and Stata (www.stata.com). Between-subject effect sizes (Cohen's *d*; Cohen, 1988) are reported for the primary outcomes.

Ethical considerations and patient and public involvement

The study was reviewed and received ethical clearance through the University of Oxford Medical Sciences Division Central University Research Ethics Committee (CUREC; Reference: R44919/RE001). Two patients with a history of sleep disturbance were independently consulted to review and comment on a version of the study protocol.

Results

Participant characteristics

Three-hundred and thirty-five individuals consented, endorsed inclusion criteria and completed online screening (Figure 1). One hundred and eleven individuals were eligible for inclusion in the study, having indicated probable “caseness” for insomnia disorder on the SCI (Espie *et al.*, 2014). Sixty-six participants responded to invitations to the laboratory and were randomised. Two did not access the sham feedback on the watch and one did not attend the second laboratory visit; hence three participants were excluded from analyses. The final sample, therefore, consisted of 63 participants (77.8% female) aged 20 to 78 years ($M = 44.38$ years, $SD = 14.45$ years; Table 1). Most identified as White British (82.5%) and were highly educated (87.3% to bachelor's degree or higher).

(Insert Table 1 about here)

Screening measures and pre-feedback sleep parameters

Scores on screening measures and sleep parameters from the night immediately preceding feedback are presented by randomised group in Table 2.

(Insert Table 2 about here)

Self-reported daytime function: First laboratory visit vs second laboratory visit

Alert cognition

ANCOVA revealed a significant effect of group on evening two, $F(1, 60) = 9.67, p = .003$, with the negative feedback group displaying significantly lower levels of *alert cognition*, relative to the positive group (estimated mean difference = -9.2, $SE = 3.0$; 95% CI [-15.1, -3.3]; $d=.79$; see Figure 3). Paired samples *t*-test confirmed a statistically significant decrease in the negative group's scores at the second laboratory visit ($M = 52.1, SE = 2.9$) when compared to the first ($M = 61.9, SE = 2.7$), $t(31) = 3.92, p < .001$. For the positive group, there was no statistically significant difference between the first ($M = 58.8, SE = 2.2$) and second laboratory visits ($M = 59.4, SE = 2.1$), $t(30) = -.33, p = .746$.

Negative mood

ANCOVA revealed a non-significant difference between groups on the second laboratory visit $F(1, 60) = 3.17, p = .080$, with the negative feedback group showing a tendency towards increased *negative mood* compared to the positive feedback group (estimated mean difference = 5.0, $SE = 2.8$; 95% CI [-0.6, 10.7]; $d=.46$).

Positive mood

ANCOVA revealed a non-significant difference between groups on the second laboratory visit, $F(1, 60) = 2.25, p = .139$ (estimated mean difference = -5.0, $SE = 3.3$; 95% CI [-11.6, 1.7]; $d=.38$).

Sleepiness/fatigue

ANCOVA revealed a significant difference between groups on the second laboratory visit, $F(1, 60) = 4.61, p = .036$, with the negative feedback group displaying increased levels of *sleepiness/fatigue*

relative to the positive group (estimated mean difference = 10.0, $SE = 4.7$; 95% CI [0.7, 19.4]; $d=.55$). A follow-up paired samples t -test confirmed that there was a statistically significant increase in *sleepiness/fatigue* in the negative group on the second laboratory visit ($M = 57.8$, $SE = 4.2$) relative to the first visit ($M = 42.8$, $SE = 3.8$), $t(31) = -5.43$, $p < .001$. For the positive group, there was no statistically significant difference between the first ($M = 48.0$, $SE = 3.2$) and second laboratory visits ($M = 50.2$, $SE = 3.0$), $t(30) = -.49$, $p = .631$.

(Insert Figure 3 about here)

Within-day trajectories in self-reported daytime function: Rise-time versus 12:00 noon and 15:00

See Table 3 for unadjusted means by group for each time-point. Linear mixed effect models showed that the positive group displayed significantly greater increase, relative to the negative group, from rise-time to 12 noon for alert cognition [adjusted mean difference (95% CI) = 8.5 (0.7, 16.3), $p=.033$] and positive mood [adjusted mean difference (95% CI) = 7.5 (1.3, 13.8), $p=.019$]. The positive group displayed greater reductions in sleepiness/fatigue from rise-time to 12 noon compared with the negative feedback group [adjusted mean difference (95% CI) = -12.2 (-22.4, -2.0), $p=.019$]. While daytime functioning tended to be poorer for the negative group at 15:00 (see Table 3) adjusted mean difference from rise-time did not reveal significant group differences for any sub-scale. No significant group differences were observed for change in negative mood at any time-point.

(Insert Table 3 here)

Sleep-related attentional bias and psychomotor vigilance

Descriptive values for the dot-probe task and PVT are presented in Table 4. With respect to sleep-related attentional bias, ANCOVA revealed no significant difference between groups on evening two, covarying for baseline scores $F(1, 60) = .62$, $p = .434$ (estimated mean difference = -3.45, $SE = 4.38$; 95% CI [-12.21, 5.31]; $d=.20$). Similarly, for the PVT, ANCOVA revealed no significant group difference

on evening two for reaction time, $F(1, 60) = .22$, $p = .638$, (estimated mean difference = -3.42, $SE = 7.24$; 95% CI [-17.91, 11.07]; $d=.12$).

(Insert Table 4 about here)

Feedback congruence, preoccupation and sleep tracking

Table 5 shows the means and standard deviations for both experimental groups on the feedback congruence and preoccupation questions administered at the end of the study protocol. Independent samples t -tests showed a statistically significant difference between groups in how much they felt the feedback was reflective of their sleep $t(61) = 3.91$, $p < .001$, with the negative group, on average, reporting feedback was more reflective of their sleep than the positive group. There was no difference on how much the groups thought about feedback during the day $t(61) = .21$, $p = .834$ – with both groups, on average, endorsing “somewhat” to the following question: *to what extent did you find yourself thinking about your sleep during the day?* No participant in either group reported using a commercial sleep monitoring device on the night of the manipulation.

(Insert Table 5 about here)

Discussion

The current study sought to test the hypothesis that giving people with insomnia sham “objective” feedback about their sleep efficiency would modify daytime functioning and sleep-related attentional processes. We specifically devised a protocol that could plausibly simulate use of wearable devices. It was predicted that relative to positive feedback, those given negative feedback would exhibit impaired indices of self-reported daytime function, higher levels of sleep-related attentional bias and lower levels of psychomotor vigilance. Congruent with hypotheses, participants given negative feedback exhibited impaired indices of subjective daytime function. Those who received negative feedback about sleep reported significantly lower levels of *alert cognition* and significantly higher levels of *sleepiness/fatigue* on the second laboratory visit. Furthermore, trajectories of daytime function

revealed lower within-day change (from rise-time to 12:00 noon) in *positive mood* and *alert cognition*, as well as elevated *sleepiness/fatigue* following negative versus positive feedback. Contrary to our hypotheses, there were no significant differences between groups for sleep-related attentional bias or psychomotor vigilance.

In the current study, two sources of evidence suggest that observed effects in subjective symptomatology may be a direct result of participants' perception of poor sleep, i.e. being given negative feedback. Within-group comparisons between laboratory visits one and two revealed an increase in symptoms for the negative group, while there was no change for the positive group. Related to this, the positive group, on average, rated their feedback as less reflective of their sleep than the negative group. This may suggest a potential ceiling effect whereby people with chronic insomnia, who by definition perceive their sleep to be poor, are less likely to believe feedback incongruent with this experience. The positive group might therefore be considered a *de facto* 'neutral' condition. Conversely, sham sleep feedback delivered to the negative group was markedly worse ($\geq 20\%$) than sleep efficiency values from both actigraphic data and sleep diary data which may have translated into a potentiation of daytime dysfunction.

Observed effects were strongest for cognition and sleepiness and were somewhat less consistent for mood variables, although trends were clearly evident and thus the study may have been underpowered to detect them. Interestingly, daytime symptom trajectories (Table 3) suggests that negative feedback may attenuate normal circadian modulation of positive mood, cognition, and sleepiness. This is in contrast to both the positive feedback group in the present study and insomnia patients from previous work, who tend to show greater diurnal variation in symptom (DISS) reports (Buysse *et al.*, 2007; Miller *et al.*, 2013). Future studies may wish to extend measurement to include other aspects of daytime functioning such as social interactions, interpersonal relationships and work productivity.

Our findings lend partial empirical support to proposed cognitive models of insomnia maintenance (Harvey, 2002; Herbert *et al.*, 2017) by showing that manipulation of sleep perception influences appraisal of daytime symptoms. However, we found no evidence that sham feedback influences sleep-related attentional bias. Whilst this may reflect a true absence of effect, there are other plausible explanations. First, the sham feedback delivered in the present study may not have been sufficiently powerful to elicit observable changes in attentional processes. This seems congruent with the absence of statistically significant differences in participants' report of daytime pre-occupation with sleep feedback. However, considering that the manipulation was powerful enough to bias subjective indices of daytime function, this explanation alone seems unlikely. Another possibility is that the task itself lacks sensitivity. While the dot-probe is sensitive to between-group comparison (patients vs controls), recent manipulation studies have failed to find evidence of treatment-related change and poor test-retest reliability (Lancee *et al.*, 2017). Indeed, we similarly observed low test-retest scores ($r = -.08$). We suggest that future studies should consider employing more nuanced and direct measures of attention allocation (e.g., eye-tracking; Harris *et al.*, 2015; Beattie *et al.*, 2017).

We similarly found no statistically significant effects for the PVT. Whilst this appears consistent with findings from the dot-probe task, it contrasts with experimental research showing changes to cognitive performance (auditory attention and processing speed) after the delivery of sham sleep feedback in students (Draganich & Erdal, 2014). Although the PVT is sensitive to the effects of basic disruption to sleep, the current findings suggest that it may not be able to detect differences that have more cognitive or expectation-based underpinnings. In future studies it will be important to examine a broader range of cognitive domains as well as other objective markers of functioning (e.g., activity levels, psychophysiological arousal).

The above conclusions are strengthened by a robust experimental design with several methodological strengths. First, we used a randomised design in which the principal experimenter (DG) was blind to group allocation, thereby minimising both selection and experimenter bias. Second, we included a

baseline night in the absence of sleep-feedback, permitting comparisons with night two, post-sham feedback. A further 'second baseline' measurement of daytime symptoms was taken immediately before delivery of feedback in the morning, enabling assessment of the effects of feedback across the day. Third, the use of an integrated actiwatch allowed for real-time assessment of daytime function, permitting verification of data capture and reducing bias from retrospective reports.

Our findings contribute to ongoing debates about the role of wearable health devices in the tracking and modification of health-related behaviours (Piwek *et al.*, 2016). Inaccurate feedback about sleep may affect hundreds of thousands of people every day, globally, driving biases in the appraisal of daytime function. While our study has focussed on those with poor sleep, such negative effects may also be relevant to normal sleepers and could be a contributory factor in the development of sleep problems. It would be prudent to test whether sham feedback may influence subsequent sleep through key cognitive mechanisms, such as potentiation of pre-sleep worry and/or sleep effort (Kyle, 2015).

The current findings must be interpreted in the context of several limitations. First, screening for sleep disorders was done using subjective report and not through formal interview or polysomnography. It is conceivable, therefore, that the sample may have included participants with comorbid sleep disorders (in addition to insomnia). Second, participants with extant psychopathology were excluded from the sample, both to mitigate issues of increased risk as a result of the manipulation and to reduce confounds which may bias results. Considering the degree to which insomnia is commonly comorbid with mental health diagnoses, this may limit generalisability. Third, the value chosen for the positive feedback condition (i.e. 91.4% sleep efficiency) was, on average, less congruent with participants' experience of sleep, which may have limited our ability to detect effects based on the influence of positive feedback. Future studies should consider tailoring feedback based on participants actual reports so as to limit gross discrepancies that may limit believability. Furthermore, we did not include a *neutral* or *no-feedback* control group limiting our ability to define what accounts for group

differences; however, the inclusion of a baseline assessment, prior to device-feedback, suggests that the observed effects may be driven by negative feedback. Clearly, a prospective study is needed to test this hypothesis directly.

Acknowledgements

We would like to thank Sam Mort of the Nuffield Department of Primary Care Health Sciences, University of Oxford, for support with statistical analysis.

References

- Altena, E., Van Der Werf, Y. D., Strijers, R. L. M. and Van Someren, E. J. W. Sleep loss affects vigilance: Effects of chronic insomnia and sleep therapy. *J. Sleep Res.* 2008, 17: 335-343.
- Bastien, C. H., Vallières, A. and Morin, C. M. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med.* 2001, 2: 297-307.
- Beattie, L., Bindemann, M., Kyle, S. D. and Biello, S. Attention to beds in natural scenes by observers with insomnia symptoms. *Behav. Res. Ther.* 2017, 92: 51-56.
- Behar, J., Roebuck, A., Domingos, J. S., Geder, E. and Clifford, G. D. A review of current sleep screening applications for smartphones. *Physiol. Meas.* 2013, 34: R29-R46.
- Buysse, D. J., Reynolds, C. F., Monk, T. H. et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989, 28: 193-213.
- Buysse, D. J., Thompson, W., Scott, J. et al. Daytime symptoms in primary insomnia: A prospective analysis using ecological momentary assessment. *Sleep Med.* 2007, 8: 198-208.
- Carney, C. E., Buysse, D. J., Ancoli-Israel S. et al. The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep.* 2012, 35: 287-302.
- Chambers, M. J. and Keller, B. Alert insomniacs: Are they really sleep deprived? *Clin. Psychol. Rev.*, 1993, 13: 649-666.
- Cohen J. *Statistical power analysis for the behavioral sciences*. Routledge Academic, New York, NY., 1988.
- Cook, J. D., Prairie, M. L. and Plante, D. T. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. *J. Affect. Disord.* 2017, 217: 299-305.

Dinges, D. F. and Powell, J. W. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods Instrum. Comput.* 1985, 17: 652-655.

Draganich, C. and Erdal, K. Placebo sleep affects cognitive functioning. *J. Exp. Psychol. Learn. Mem. Cogn.* 2014, 40: 857-864.

Espie, C. A., Kyle, S. D., Hames, P., Gardani, M., Fleming, L. and Cape, J. The Sleep Condition Indicator: a clinical screening tool to evaluate insomnia disorder. *BMJ Open.* 2014, 4: e004183.

Espie, C. A., Kyle, S. D., Williams, C. et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. *Sleep.* 2012,35: 769-781.

Faul, F., Erdfelder, E., Lang, A-G. and Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* 2007, 39: 175-191.

Fortier-Brochu, É., Beaulieu-Bonneau, S., Ivers, H. and Morin, C. M. Insomnia and daytime cognitive performance: A meta-analysis. *Sleep Med. Rev.*, 2012, 16: 83-94.

Harris, K., Spiegelhalter, K., Espie, C. A., MacMahon, K. M. A., Cleland, H. and Kyle, S. D. Sleep-related attentional bias in insomnia: A state-of-the-science review. *Clin. Psychol. Rev.* 2015, 42: 16-27.

Harvey, A. G. A cognitive model of insomnia. *Behav. Res. Ther.* 2002, 40: 869-893.

Harvey, A. G. and Tang, N. (Mis)perception of sleep in insomnia: A puzzle and a resolution. *Psychol. Bull.*, 2012, 138: 77-101.

Herbert, V., Pratt, D., Emsley, R. and Kyle, S. D. Predictors of nightly subjective-objective sleep discrepancy in poor sleepers over a seven-day period. *Brain Sci.* 2017, 7: 29-44.

Ko, P. R. T., Kientz, J. A., Choe, E. K., Kay, M., Landis, C. A. and Watson, N. F. Consumer sleep technologies: A review of the landscape. *J. Clin. Sleep Med.* 2015, 11: 1455-1461.

Kroenke, K., Spitzer, R. L. and Williams, J. B. W. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* 2001, 16: 606-613.

Kyle, S. D. Sleep supervision. In: P. Lewis and R. A. Page (Eds) *Spindles: Stories from the science of sleep*. Comma Press, Manchester, 2015: 112-117.

Lancee, J., Yasin, S. L., Brendel, R. S., Boffo, M., Clarke, P. J. F. and Saleem, E. Attentional bias modification training for insomnia: A double-blind placebo controlled randomized trial. *PLoS One*. 2017, 12: e0174531.

Lee, J. and Finkelstein, J. Activity trackers: A critical review. *Stud. Health. Technol. Inform.* 2014, 205: 558-562.

Manea, L., Gilbody, S. and McMillan, D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ*, 2012, 184: E191-6.

MacMahon, K. M. A., Broomfield, N. M. and Espie, C. A. Attention bias for sleep-related stimuli in primary insomnia and delayed sleep phase syndrome using the dot-probe task. *Sleep*. 2006, 29: 1420-1427.

Miller, C. B., Kyle, S. D., Marshall, N. S. and Espie, C. A. Ecological momentary assessment of daytime symptoms during sleep restriction therapy for insomnia. *J. Sleep Res.* 2013, 22: 266-272.

Mulligan, L. D., Haddock, G., Emsley, R., Neil, S. T. and Kyle, S. D. High resolution examination of the role of sleep disturbance in predicting functioning and psychotic symptoms in schizophrenia: A novel experience sampling study. *J. Abnorm. Psychol.* 2016, 125: 788-797.

Netzer, N. C., Stoohs, R. A., Netzer, C. M., Clark, K. and Strohl, K. P. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann. Intern. Med.* 1999, 131: 485-491.

Olsen, M. K., Stechuchak, K. M., Edinger, J. D., Ulmer, C. S. and Woolson, R. F. Move over LOCF: Principled methods for handling missing data in sleep disorder trials. *Sleep Med.* 2012, 13: 123-132.

Patel, M. S., Asch, D. A. and Volpp, K. G. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA.* 2015, 313: 459-460.

Piwek L., Ellis, D. A., Andrews, S. and Joinson, A. The rise of consumer health wearables: Promises and barriers. *PLoS Med.* 2016, 13: 1-9.

Richardson, C., Gradisar, M. and Pulford, A. The development of insomnia or the plasticity of good sleep? A preliminary study of acute changes in sleep and insomnia resulting from an analogue trauma. *Behav. Sleep Med.* 2014, 12: 37-41.

Russell, C., Wearden, A. J., Fairclough, G., Emsley, R. A. and Kyle, S. D. Subjective but not actigraphy-defined sleep predicts next-day fatigue in chronic fatigue syndrome: A prospective daily diary study. *Sleep.* 2016, 39: 937-944.

Semler, C. N. and Harvey, A. G. Misperception of sleep can adversely affect daytime functioning in insomnia. *Behav. Res. Ther.*, 2005, 43: 843-856.

Spitzer, R. L., Kroenke, K., Williams, J. B. W. and Lo, B. A brief measure for assessing generalized anxiety disorder. *Arch. Intern. Med.* 2006, 166: 1092-1097.

Van den Bulck, J. Sleep apps and the quantified self: blessing or curse? *J. Sleep Res.* 2015, 24: 121-123.

Wilson, S. J., Nutt, D. J., Alford, C. et al. British Association for Psychopharmacology consensus statement on evidence-based treatment of insomnia, parasomnias and circadian rhythm disorders. *J Psychopharmacol.* 2010, 24: 1577-1601.