UNIVERSITY of York

This is a repository copy of *Development And Early Validation Of A Patient Reported Outcome Measure To Assess Sleep Amongst People Experiencing Problems With Alcohol Or Other Drugs.*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/126144/

Article:
Neale, Joanne, Vitoratou, Sillia, Lennon, Paul et al. (4 more authors) (2018) Development And Early Validation Of A Patient Reported Outcome Measure To Assess Sleep Amongst People Experiencing Problems With Alcohol Or Other Drugs. Sleep. ISSN 1550-9109

https://doi.org/10.1093/sleep/zsy013

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

SLEEPJ, 2018, 1–14

doi: 10.1093/sleep/zsy013
Advance Access Publication Date: 10 January 2018
Original Article

ORIGINAL ARTICLE

Development and early validation of a patient-reported outcome measure to assess sleep amongst people experiencing problems with alcohol or other drugs

Joanne Neale,[1,*] Silia Vitoratou,[2] Paul Lennon,[3] Robert Meadows,[4] Sarah Nettleton,[5] Daria Panebianco,[1] John Strang[1] and John Marsden[1]

[1]National Addiction Centre, 4 Windsor Walk, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, [2]Department of Biostatistics and Health Informatics, Psychometrics and Measurement Lab, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London, UK, [3]Aurora Project and Service User Research Group, London, UK, [4]Department of Sociology, University of Surrey, Guildford, Surrey, UK and [5]Department of Sociology, University of York, Heslington, York, UK
Work Performed: King's College London, UK

*Corresponding author. Joanne Neale, National Addiction Centre, 4 Windsor Walk, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, London, SE5 8AF, UK. Email: joanne.neale@kcl.ac.uk.

Abstract

Study Objectives: To develop a patient-reported outcome measure to assess sleep amongst people experiencing problems with alcohol or other drugs.

Methods: Item development included secondary analyses of qualitative interviews with drug or alcohol users in residential treatment, a review of validated sleep measures, focus groups with drug or alcohol users in residential treatment, and feedback from drug or alcohol users recruited from community and residential settings. An initial version of the measure was completed by 549 current and former drug or alcohol users (442 in person and 107 online). Analyses comprised classical test theory methods, exploratory and confirmatory factor analysis, measurement invariance assessment, and item response theory (IRT).

Results: The initial measure (30 items) had good content and face validity and was named the Substance Use Sleep Scale (SUSS) by addiction service users. After seven items were removed due to low item-factor loadings, two factors were retained and labeled: "Mind and Body Sleep Problems" (14 items) and "Substance-Related Sleep Problems" (nine items). Measurement invariance was confirmed with respect to gender, age, and administration format. IRT (information) and classical test theory (internal consistency and stability) indicated measure reliability. Standard parametric and nonparametric techniques supported convergent and discriminant validity.

Conclusions: SUSS is an easy-to-complete patient–reported outcome measure of sleep for people with drug or alcohol problems. It can be used by those concerned about their own sleep, and by treatment providers and researchers seeking to better understand, assess, and potentially treat sleep difficulties amongst this population. Further validity testing with larger and more diverse samples is now required.

Statement of Significance

The Substance Use Sleep Scale (SUSS) is the first sleep measure designed specifically for people experiencing problems with alcohol or other drugs. It comprises 23 items and two factors: "Mind and Body Sleep Problems" and "Substance-Related Sleep Problems." SUSS was developed with significant input from substance users and can be used by them to monitor and reflect on their own sleep; by treatment providers to encourage and enable people who use substances to think about sleep and identify strategies for improving sleep; and by researchers and others as an outcome measure when designing and implementing sleep interventions for this population. Further validity testing, involving larger and more diverse samples, is now needed.

Key words: sleep; measurement; PROM; drugs; alcohol; addiction treatment

Submitted: 8 July, 2017; **Revised:** 4 December, 2017
© Sleep Research Society 2018. Published by Oxford University Press [on behalf of the Sleep Research Society].
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

Introduction

In a comprehensive review of the biomedical literature on sleep and substance use disorders, Arnedt et al. report that "nearly all substances ingested prior to bedtime alter the subjective and objective experience of sleep" (p.527) [1]. For example, the acute administration of drugs in the opioid class can increase nocturnal arousals and stage shifts, decrease total sleep time and sleep efficiency, and reduce the time spent in rapid eye movement (REM) sleep. Alcohol suppresses REM sleep in healthy individuals and is associated with short sleep duration in people who are alcohol dependent. Additionally, people experiencing sleeping difficulties often consume drugs or alcohol to help them fall asleep—thus indicating that sleep problems are a risk factor for substance misuse [1].

The sociological literature on sleep and substance use disorders is less well developed than the biomedical literature, although complex interactions between sleep, substance use, treatment, and recovery processes have been reported. In one qualitative study of current and former heroin users, participants repeatedly complained that sleeping problems caused them to feel distressed, exhausted, and unable to cope [2, 3]. Sleep tended to worsen during periods of detoxing, with individuals reporting difficulty falling asleep, waking throughout the night, waking early in the morning, having "restless," "jumpy," and "twitchy" legs, and vivid dreaming (including nightmares and disturbing dreams with drug-using content). Difficulties sleeping and tiredness constrained their capacity to participate in therapeutic activities when in treatment, had the potential to trigger relapse, and generally undermined recovery efforts [2, 3].

In another qualitative study (comprising focus groups with people reporting a drink or drug problem, being treated in residential detoxification or rehabilitation, or defining themselves as abstinent), "sleeping well" was consistently identified as an important indicator of recovery from addiction [4]. This finding supported the earlier study of current and former heroin users in which individuals who were attempting to reduce or abstain from heroin use described how they tried hard to establish more regular sleeping patterns, explained that they often found the structured bedtime routines of residential drug treatment helpful, and expressed relief and satisfaction when sleep patterns started to improve [3].

Within the general population, the relationship between sleep and health (physical and mental) has been widely documented [5, 6]. Sleep deficiency is associated with anxiety and depression, weight gain, impaired immune response, and increased risk of numerous diseases [5–11]. Tiredness decreases motor and cognitive performance, reducing communication and decision-making, impairing memory and concentration, and increasing the risk of accidents and injuries [5, 6, 12, 13]. Sleep duration has also been linked to all-cause mortality in a number of populations [5, 6, 14, 15]. Given the importance of sleep to health, scientists and clinicians have developed ways of evaluating sleep and sleep-related impairments [5]. These include functional imaging and electrophysiological techniques and actigraphy. However, the most practical and widely used tools for evaluating sleep are self-report instruments, for example, sleep diaries, but also questionnaires that have been validated to assess particular types of sleep disturbance or to characterize symptoms of particular sleep disorders [5].

Although validated sleep measures have been used in a number of studies of people addicted to alcohol or other drugs [16–22], there is currently no sleep measure developed explicitly for people experiencing substance dependence. This is a limitation given that individuals who report problems with alcohol or other drugs may experience particular constellations of sleep-related difficulties, which change depending on the substances taken, whether an individual is detoxing or not, and whether an individual has sleep routines imposed on them (e.g. in residential treatment or prison). The lack of a specific standardized, validated questionnaire makes it difficult for those experiencing substance-related impaired sleep to convey the nature and strength of their sleep problems. It also poses challenges for those seeking to better understand, assess, and potentially treat sleep difficulties amongst alcohol and other drug users.

Self-completion questionnaires that assess subjective health status are often referred to as patient-reported outcome measures (PROMs) [23]. PROMs are used across many areas of medicine, but have received comparatively less attention from within the drug and alcohol sector, perhaps reflecting the fact that historically alcohol and other drug users have not tended to be widely consulted by treatment providers or scale developers. Although the methodology for developing PROMs is constantly evolving, it is generally agreed that individuals from the target patient population should be involved in item and scale development. This helps us to ensure that the constructs measured and the language and terminology used are acceptable to, and reflect the priorities and preferences of, those who will later complete the measure [4, 24, 25]. Once developed, PROMs should be subjected to rigorous psychometric testing [23].

The aim of our study was to develop a new PROM to assess sleep amongst people experiencing problems with alcohol or other drugs. The work was undertaken in two phases. Phase 1 identified items for the new PROM (ensuring good face and content validity, acceptability, and usability for the target population). Phase 2 then evaluated the psychometric properties and factorial structure of the new PROM. The study received ethical approval from a University Research Ethics Committee.

Methods

Phase 1 (hereafter, "item development") occurred in three stages between July and October 2015. Data collection for phase 2 (hereafter, "measurement evaluation") occurred in two subsequent stages between November 2015 and August 2016. To be eligible to participate in the study, individuals needed to (1) be over 18 years of age, (2) self-report current or previous problem substance use (illicit drugs or alcohol), (3) have sufficient understanding of English to be able to complete a basic questionnaire alone or with reading support, and (4) be able to give informed consent.

Support and advice were provided to the research team by two separate Project Advisory Groups (PAGs). The first PAG (PAG 1) comprised addiction service users, addiction clinicians and researchers, sleep clinicians and researchers, and a PROM expert (n = 10); it met once during November 2015. The second PAG (PAG 2) comprised addiction service users only (n = 11); this group met several times throughout the study and also provided ad hoc advice on a small group and individual basis by telephone, email, and in person (IP) as issues needing discussion arose.

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

Service user advisors were each paid £20 per consultation; other advisors were not paid.

Item development

Stage 1: Identifying candidate items

Stage 1 involved secondary analyses of a pre-existing qualitative data set generated during a separate but linked study of sleep and addiction in residential rehabilitation settings [26, 27]. This prior study had been undertaken in residential treatment services in England during 2014 and 2015. Semi-structured interviews had been conducted with 19 women and 9 men (n = 28). Their ages had ranged from 24 to 83 years, and they had self-reported a mixture of alcohol, illicit drug, and prescription drug misuse. For the present study, the transcribed interview data were reanalyzed to identify sleep-related themes that might constitute candidate items for the new PROM.

In addition, the research team separately reviewed twelve validated sleep measures that had all previously been used in studies of people addicted to drugs or alcohol (Supplementary Table S1). Items that complemented or supported the themes discussed in the 28 qualitative interviews were added to the list of candidate PROM items. The list was then screened by four addiction service users (2 men and 2 women) from PAG 2. The four PAG 2 members were asked to suggest additional items, reword items, or remove irrelevant items. They were also consulted on the most appropriate time scale and scoring system for the new PROM.

Stage 2: Creating a draft measure

In stage 2, all the candidate items identified in stage 1 were discussed within two focus groups of treatment clients. Focus group participants were recruited from two residential services, both of which provided detoxification and structured support (neither service had been involved in stage 1). Participants included 9 men and 3 women (n = 12), their ages ranged from 27 to 47 years, and they reported misuse of alcohol, illicit drugs, and prescribed drugs (often in combination). Each focus group participant received a £15 shopping voucher in compensation for their time.

The focus group participants were invited to debate the long list of candidate items from stage 1, add new items, suggest any changes to wording, or remove any items. Feedback from the focus groups was discussed with members of PAG 2 on two separate occasions and the long list of items was revised according to their advice. Following this, a draft sleep PROM was prepared by the research team.

Stage 3: Assessment of face and content validity, acceptability, and usability

In stage 3, a new sample of 30 current and former drug and alcohol users completed the draft PROM in person, commenting on the content, time scale, wording, scoring system, and layout. Participants were recruited through a variety of community settings (n = 22), residential settings (n = 4), and an outreach setting for people sleeping on the streets (n = 4). They included 26 men and 4 women; age range = 29–65 years. In total, 27 had used drugs or alcohol in the last 6 months. Of these, 10 reported that their main substance was an illicit drug, 8 reported that their main substance was alcohol, and 9 reported that they used illicit drugs and alcohol equally. All participants received a £15 voucher in compensation for their time. The research team used the participants' feedback to modify the draft sleep measure, and the modified version was then discussed with PAG 1 members.

Measurement evaluation

Stage 4: In Person (IP) sample

For stage 4, we recruited current and recent drug and alcohol users (n = 442) from community treatment services, homeless hostels, and peer support services across five English towns and cities. These individuals completed a questionnaire that comprised basic demographic, drug use and sleep questions, and the draft sleep measure. To maximize recruitment and completion, the questionnaire was kept as short as possible so that it could be answered relatively quickly (< 15 min), even by people who might have limited literacy or find it difficult to concentrate due to drug or alcohol withdrawal symptoms. Participants were offered refreshments to compensate for their time.

In order to conduct more advanced validation work, additional data (including some follow-up data) were required. As we were concerned about the burden, this would place on our target population, and also the difficulty of tracking and relocating participants across many locations simultaneously, we only collected these additional data from a subsample of participants attending services in one city. We recruited to preset targets that we believed would be both feasible and adequate for the analyses and front-ended the data collection to ensure that our targets were met. Accordingly, the first 100 participants also completed two validated measures: (1) the Pittsburgh Sleep Quality Index (PSQI) [28] and (2) the Substance Use Recovery Evaluator (SURE) [29]. The PSQI is a self-rated questionnaire which assesses sleep quality and disturbances over a 1 month time interval, score range 0–21 where lower scores denote better sleep quality. SURE is a psychometrically valid PROM for recovery from drug and alcohol dependence, score range 21–63 where higher scores denote greater recovery. Of these 100 individuals, the first 42 completed the questionnaire and validated measures a second time, 2–5 days later. These 42 individuals received a £10 supermarket voucher for each questionnaire completed. Of these 42 individuals, 22 also wore actigraphs for a period of 7 days. These 22 participants received a further £10 voucher for the inconvenience of wearing the watches.

Stage 5: Online (OL) sample

To expand the geographical reach of the data collection and to ensure that we recruited beyond current treatment populations, an online (OL) version of the demographic, drug use and sleep questions and the sleep measure was created using the survey tool BOS (https://www.onlinesurveys.ac.uk/). The OL survey was open to anyone who was currently in, or who had previously been in, community or residential treatment for an alcohol or other drug problem. The survey link was circulated to service user organisations and treatment services via social media (Twitter, LinkedIn) and email. No compensation was offered for completing the survey. In total, 107 individuals responded OL.

Statistical analysis

Actigraph data were entered into the software programme Sleep Analysis 7, and estimates of (1) sleep efficiency, (2) sleep latency,

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

(3) total sleep time, and (4) wake bouts for each participant were created. We used parametric (t test, Pearson's correlation coefficient) and nonparametric methods (Mann–Whitney test, Spearman's correlation coefficient) to test score differences and associations, subject to the symmetry of the distributions.

To assess the reliability of the new measure, both classical test theory (internal consistency and stability) and item response theory (IRT; information) approaches were used. Internal consistency was evaluated via Cronbach's α coefficient [30], along with the item–total correlations and the computation of α if the item was omitted. Stability was evaluated via Cohen's κ for each item [31], following Landis and Koch interpretations [32], along with the percentage of agreement. For the total scores (continuous variables), a (two-way mixed) intraclass correlation coefficient [33] was also calculated.

We performed item factor analysis (IFA) [34] suitable for binary items to identify the dimensionality of the measure. Both exploratory (EFA) and confirmatory (CFA) models were used, in different samples. Measurement bias (or violation of measurement invariance) was evaluated via multiple groups of CFA for categorical exogenous variables (group membership) [35–37]. This was conducted in three stages. First, metric invariance was tested by restricting the loading of each item to its corresponding factor to be equal across groups. Each item that demonstrated metric invariance was equivalently related to its factor across groups. Second, scalar invariance was tested by restricting the thresholds of each item so that they were equal across groups. Each item that showed scalar invariance had the same probability of a positive response for individuals of the same trait level across groups. Third, strict invariance was tested by restricting the residual variances of each item so that they were equal across groups. Each item that demonstrated strict invariance was explained similarly by their factor across groups. For continuous exogenous variables, such as age, measurement invariance was tested using the Multiple Indicators, Multiple Causes (MIMIC) model [38, 39]. A significant direct effect from the exogenous variable to a particular item demonstrated a lack of invariance, so raising concerns of measurement bias.

Classical test theory assesses the whole test as the unit of measurement, whereas IRT examines individual items or questions. IRT is a probabilistic model of the mathematical relationship between individuals' abilities (or other hypothesized traits) and the item characteristics. In other words, the probability that an individual will respond positively or correctly to an item designed to measure a particular trait is a function of the item's difficulty, the item's discrimination ability, and the amount of the underlying trait possessed by the individual. The two-parameter IRT model (2PL-IRT) [40] was used to evaluate the severity (difficulty), discrimination ability, and information (precision) of the items within each dimension of the final measure. Evidence of (concurrent) convergent and divergent validity was assessed via standard parametric and nonparametric techniques, as described above. The statistical software Mplus (version 7) was used for all latent trait models (IRT, IFA, EFA, CFA, and MIMIC).

The statistical software Mplus (version7) was used for all latent trait models (IRT, IFA, EFA, CFA, and MIMIC) and the following goodness of fit measures are reported: the relative chi-square (rel $\chi 2$: with preferred values close to 2; Hoelter [41]), the Root Mean Square Error of Approximation (RMSEA, with preferred values less than 0.8; Browne and Cudeck [42]), the Taylor–Lewis Index (TLI, with preferred values higher than 0.9; Bentler and Bonett [43]), and the Comparative Fit Index (CFI, with preferred values higher than 0.9; Bentler [44]).

Results

Item development outcome

Stage 1: Identifying candidate items

Secondary analyses of the 28 qualitative interviews identified 72 themes for consideration as candidate PROM items (Supplementary Table 2). Following review of the 12 validated sleep measures, 21 new themes were added to the list of candidate items for the new sleep PROM (Supplementary Table 3). In consultation with the four PAG 2 members, the long list of candidate items was turned into 75 simple "no"/"yes" statements relating to sleep in the previous week (Supplementary Table 4). Binary response options were preferred over polytomous response options because they were considered easier for respondents to understand and complete. The last week was preferred over shorter and longer time frames as this was deemed long enough to capture sleep patterns without responses being undermined by poor recall. Supporting this, there is good evidence that other sleep-related PROMs, such as the PSQI and the Epworth Sleep Scale, can be administered with confidence for week-long reporting periods for between-subject analyses [45]. The expression "last week" was preferred to "last 7 days" since the word "days" could generate confusion amongst a population that has a tendency to sleep during the daytime (rather than at night) when using substances but not when in treatment or abstinent [2, 3].

Stage 2: Creating a draft measure

Feedback from the focus group participants and PAG 2 members indicated that some of the 75 sleep statements needed to be amended or removed. A key problem was that many focused on sleep "behaviors" rather than sleep "problems," and these indicators of sleep behavior were not necessarily indicative of sleep problems. For example, discussions revealed that "sleeping with the television or radio or music on" or "drinking a hot drink before bed," or "napping during the day" could help some people to sleep, impede the sleep of others, and have no effect on yet others. Items considered ambiguous by the PAG were therefore removed.

The resultant draft PROM comprised 30 × "no"/"yes" statements (scoring "0" and "1"; a higher score denoting more sleep problems). These statements focused on the previous week and related to seven broad aspects of sleep: "sleep satisfaction," "sleep environment," "falling asleep," "night time activity," "sleep quality," "waking up," and "daytime functioning" (Supplementary Table 5).

Stage 3: Assessment of face and content validity, acceptability, and usability

In the final development stage, 30 individuals completed the draft PROM. In response to a series of structured questions, all 30 reported that the PROM was easy to understand, all 30 reported that it was easy to complete, 27 reported that the length was about right (two said it was too short and one said it was too long), and 20 reported that they had enjoyed completing it (10 felt neutral and nobody actively disliked it). In total, 28/30 said

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

that there were no irrelevant questions and 28/30 thought the PROM covered everything necessary. Completion times ranged from 2 to 20 min (mean 9 min; with those taking longer often discussing their responses with the researcher because they seemed to be finding the questionnaire interesting). Following consultations with members of PAG 1, only a small number of minor wording changes were made, and good face and content validity, acceptability, and usability were confirmed. Members of PAG 2 named the measure the Substance Use Sleep Scale (SUSS).

Measurement evaluation outcome

Sample characteristics

Several differences between the demographic characteristics of the individuals who completed the measure IP and those who completed OL were evident (Table 1). The OL sample had higher percentages of women (63.6% vs 30.5%), individuals of white ethnicity (95.3% vs 78.7%), and individuals who left school after the age of 16 (77.6% vs 31.4%). Age did not differ significantly between the two samples, with the total sample being between 20 and 71 years old. There were no differences with respect to length of problem substance use, and 9% of the total sample reported less than 10 years of drug or alcohol problems.

For the exploratory and confirmatory factor analysis, we used a random number algorithm (automatically generated by the software SPSS) to divide the IP sample (n = 442) into two split halves (hereafter, IP-a and IP-b). Randomization produced two split halves that did not differ in terms of demographic or clinical characteristics. As the OL sample (n = 107) was too small to be similarly divided, it was instead used as a second confirmation sample to strengthen the analyses.

Item selection

We began our analyses by testing for potential problematic items, using classical test theory tools. Grouping all items together, Cronbach's α coefficient was 0.88 for the complete sample (IP: 0.88; OL: 0.85). In the IP sample, there was no improvement in the reliability index by omitting items, and the item–total correlations (biserial correlation coefficient r) varied between r = 0.3 and r = 0.6. In the OL sample, there were some low item–total correlations (varying between 0.1 and 0.6), but no substantial reduction in α when items were omitted (Table 2). According to Landis and Koch [32], the level of agreement was fair to perfect (Cohen's κ varied between 0.4 and 1). Agreement was higher than 73% in all cases, even for lower values of κ. Therefore, in terms of reliability (according to classical test theory assumptions), no problematic items were found and so all items were used in the EFA.

Exploratory IFA

All 30 items were included in an EFA model (sample IP-a). The sample correlation matrix produced seven eigenvalues larger than 1 (12.8, 3.1, 1.8, 1.6, 1.4, 1.1, 1.1). According to Kaiser's criterion, this suggested up to seven extracted factors. The scree plot, however, indicated a two-factor solution (Figure 1).

Table 3 presents the goodness of fit indices for all models, from the unidimensional solution to the seven-factor solution. Close fit was achieved at the two-factor solution. Increasing the number of factors to three or more resulted in nonsalient loadings, cross-loadings, and/or factors with a small number of items (and thus small reliability). All seven solutions were evaluated in terms of the content of the extracted dimensions, and the two-factor solution was considered the most satisfactory in terms of face validity.

Three items ("disturbed by noise," "disturbed by light," and "woken up in the night and drunk caffeinated and/ or sugary drinks") had loadings equal to or less than 0.3 on both factors, so they were omitted and the analysis was repeated. Four more items ("sleeping tablets or medicines to help me sleep," "woken up short of breath," "needed caffeine and/ or sugary foods or drink to get through the day," "needed to sleep or to nap during the day") were then also omitted since they had smaller loadings

Table 1. Descriptive indices by sample and in total

	In person (N = 442)		Online (N = 107)		Total (N = 549)		
	N	%	N	%	N	%	Comparison
Age left school							
≤16 years	301	68.1	24	22.4	325	59.2	χ^2 = 18.568, df = 1, p < 0.001
>16 years	139	31.4	83	77.6	222	40.4	
Ethnicity							
White	348	78.7	102	95.3	450	82.0	χ^2 = 16.049, df = 1, p < 0.001
Other	94	21.3	5	4.7	99	18.0	
Gender							
Female	135	30.5	68	63.6	203	37.0	χ^2 = 40.277, df = 1, p < 0.001
Male	307	69.5	39	36.4	346	63.0	
	Mean (*SD*)	Median (min–max)	Mean (*SD*)	Median (min–max)	Mean (*SD*)	Median (min–max)	Comparison
Age (years)	44.3 (9.8)	45 (20–69)	45.5 (9.7)	46 (22–71)	44.6 (9.8)	45 (20–71)	t = 1.144, df = 545, p = 0.253
Duration of substance use (years)	20.1 (11.1)	20 (0–47)	19.9 (10.3)	20 (0–50)	20 (10.9)	20 (0–50)	t = –0.212, df = 543, p = 0.833
PSQI score	9.8 (4.2)	9 (1–18)					
SURE score	48.1 (9.7)	47 (26–63)					

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

Table 2. Reliability: internal consistency indices by sample and test–retest agreement

	IP sample (*N* = 442)		OL sample (*N* = 107)		Test–retest (*N* = 42)	
Statement	Corrected item–total correlation	Cronbach's α if item deleted	Corrected item–total correlation	Cronbach's α if item deleted	Cohen's κ (95% CI)	% of agreement
Wanted to sleep better	0.45	0.88	0.61	0.84	0.4 (0.0, 0.7)	78.6
Worried about my sleeping	0.48	0.88	0.52	0.84	0.6 (0.4, 0.9)	81.0
Woken up tired most mornings	0.49	0.88	0.54	0.84	0.8 (0.6, 1.0)	92.9
Disturbed by noise	0.31	0.88	0.26	0.85	0.6 (0.3, 0.8)	78.6
Disturbed by light	0.31	0.88	0.23	0.85	0.5 (0.2, 0.9)	85.7
Felt too unsafe to sleep	0.36	0.88	0.41	0.84	0.4 (0.0, 0.9)	90.5
Difficulty falling asleep	0.55	0.88	0.44	0.84	0.9 (0.7, 1.0)	92.9
Uncontrollable/racing thoughts when I tried to sleep	0.56	0.88	0.54	0.84	0.8 (0.6, 1.0)	90.5
Negative emotions when I tried to sleep	0.57	0.88	0.53	0.84	0.8 (0.6, 1.0)	88.1
Sleeping tablets or medicines to help me sleep	0.35	0.88	0.14	0.85	0.8 (0.6, 1.0)	90.5
Drunk alcohol to help me sleep	0.41	0.88	0.40	0.84	0.8 (0.6, 1.0)	95.2
Street drugs to help me sleep	0.40	0.88	0.16	0.85	0.7 (0.4, 1.0)	90.5
Waking up lots in the night	0.43	0.88	0.39	0.84	0.8 (0.6, 1.0)	90.5
Woken up in the night and drunk caffeinated and/or sugary drinks	0.21	0.88	0.17	0.85	0.7 (0.4, 0.9)	88.1
Woken up in the night and drunk alcohol	0.37	0.88	0.30	0.85	0.9 (0.7, 1.0)	97.6
Woken up in the night and used street drugs	0.41	0.88	0.31	0.85	0.5 (0.0, 0.9)	90.5
Woken up in the night and smoked tobacco	0.40	0.88	0.37	0.84	1.0 (0.9, 1.0)	97.6
Panic attacks in the night	0.46	0.88	0.29	0.85	0.6 (0.3, 0.9)	88.1
Vomited in my sleep	0.26	0.88	-	-	1	100
Dreams which disturbed sleep	0.42	0.88	0.35	0.84	0.6 (0.3, 0.8)	81.0
Aches and pains that stopped me from sleeping	0.45	0.88	0.44	0.84	0.6 (0.4, 0.9)	81.0
Felt restless in my sleep	0.53	0.88	0.49	0.84	0.7 (0.4, 0.9)	83.3
Woken up feeling confused or disoriented	0.49	0.88	0.34	0.85	0.7 (0.4, 0.9)	85.7
Woken up with a hangover or drunk	0.41	0.88	0.39	0.84	0.7 (0.4, 1.0)	95.2
Woken up withdrawing	0.42	0.88	0.43	0.84	0.7 (0.4, 1.0)	90.5
Needed alcohol or drugs to get out of bed	0.46	0.88	0.31	0.85	0.4 (0.0, 0.8)	88.1
Woken up short of breath	0.46	0.88	0.24	0.85	0.5 (0.2, 0.9)	85.7
Too tired to think clearly or to do things during the day	0.58	0.88	0.53	0.84	0.5 (0.2, 0.7)	73.8
Needed caffeine and/ or sugary foods or drink to get through the day	0.31	0.88	0.29	0.85	0.7 (0.5, 0.9)	85.7
Needed to sleep or to nap during the day	0.35	0.88	0.32	0.85	0.8 (0.5, 1.0)	88.1

compared with the rest of the items within their factor (<0.6 in all cases). This was undertaken as a stepwise procedure, omitting one item at the time (but only where omitting the item increased the content validity of the corresponding factors, as judged by the research team). The two-factor model was then refitted for the remaining items. The process stopped when (1) all items loaded at least 0.6, (2) the goodness of fit measures indicated a robust solution, and (3) team members confirmed that the content of each factor was adequately described by the items. The final solution had close fit to the data: rel $\chi^2 = 1.66$, RMSEA = 0.055 with 95% CI: (0.04, 0.07), CFI = 0.95, and TLI = 0.94. The corresponding EFA loadings, under OBLIMIN rotation, are presented in Table 4.

Confirmatory IFA

The two-factor (23-item) solution also had a close fit to the IP-b data: rel $\chi^2 = 1.82$, RMSEA = 0.061 with 95% CI: (0.05, 0.07), CFI = 0.94, and TLI = 0.94. Additionally, the values of the goodness

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

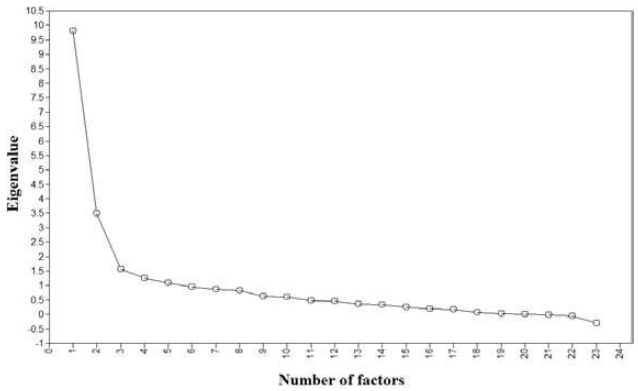

Figure 1. Scree plot (sample OL-a).

of fit indices and the loadings resembled those in EFA (Table 4). CFA was next applied to the OL sample data, revealing a close fit (rel χ^2 = 1.31, RMSEA = 0.054 with 95% CI: (0.03, 0.07), CFI = 0.92, and TLI = 0.92) and large loadings.

The two factors were named based on the content of their items, with factor 1 referred to hereafter as "Mind and Body Sleep Problems" (MBSP) and the second factor referred to as "Substance-Related Sleep Problems" (SRSP).

Measurement invariance

The next stage of our analysis involved testing for potential measurement bias in relation to how the data were collected (IP or OL), age, and gender. For this, the two random split halves were combined and the total IP sample was used.

Invariance in IP and OL samples

Metric invariance held between the OL and IP samples (DIFFTEST: χ^2 = 30.51, *df* = 20, *p* = 0.06). Full scalar invariance did not hold (DIFFTEST: χ^2 = 54.83, *df* = 20, *p* < 0.001) and the modification indices suggested that the fit of the model could be improved by allowing the thresholds of the items "drunk alcohol to help me sleep" and "woken up with a hangover or drunk" to vary. By allowing this, partial scalar invariance was held (DIFFTEST: χ^2 = 22.42, *df* = 18, *p* = 0.214). When evaluating the residual variances, strict invariance also held (DIFFTEST: χ^2 = 30.68, *df* = 22, *p* = 0.103).

All items related to their factor (MBSP or SRSP) similarly across the IP and OL samples (metric invariance). For 21/23 items, individuals who completed the questionnaire IP had the same expected response as individuals who completed the questionnaire OL, when they had the same levels of MBSP and SRSP (partial scalar invariance). However, for the same levels of SRSP, the IP sample had a greater probability of reporting "drinking alcohol to help sleep" (threshold: IP = 0.68, OL = –0.45) and a greater probability of reporting "waking up with a hangover" (threshold: IP = 1.1, OL = –0.32). The amount of item variance explained by each factor was the same across the IP and OL samples for all items (strict invariance).

As there was full invariance with respect to loadings and residuals and 91% invariance with respect to the thresholds, we concluded that there was no substantial measurement bias, and the IP and OL samples were merged for the remaining analyses.

Table 3. Goodness of fit indices—EFA-sample IP-a (N = 221)

# of factors	Rel χ^2	RMSEA	CFI	TLI
One	2.1	0.071	0.89	0.88
Two	1.4	0.042	0.96	0.96
Three	1.2	0.032	0.98	0.98
Four	1.1	0.026	0.99	0.99
Five	1.1	0.015	>0.99	>0.99
Six	1.1	0.013	>0.99	>0.99
Seven	1.1	0.004	>0.99	>0.99

Invariance with respect to gender and age

When the IP and OL data were combined for the gender and age invariance tests, full measurement invariance held with respect to gender (Metric DIFFTEST: χ^2 = 19.72, *df* = 21, *p* = 0.539; Scalar DIFFTEST: χ^2 = 29.75, *df* = 21, *p* = 0.097; Strict DIFFTEST: χ^2 = 21.74, *df* = 23, *p* = 0.536).

We then used the MIMIC model to detect direct effects of age (measured on a continuous scale) on individual items. Assuming the same levels of MBSP, increasing age was associated with reduced probability of people reporting both "difficulty falling asleep" (d.e. = –0.025, *p* = 0.001) and "negative emotions when trying to sleep" (d.e. = –0.023, *p* = 0.004). In contrast, for the same levels of MBSP, increasing age was associated with increased probability of people reporting "aches and pains that stopped sleeping" (d.e. = 0.044, *p* < 0.001). For the same levels of SRSP, the probability of reporting "drinking alcohol to help sleep" (d.e. = 0.024, *p* = 0.018) and "waking up in the night and drinking alcohol" (d.e. = 0.044, *p* < 0.001) both increased with age. However, the magnitude of the direct effects was very small and our results therefore indicated that age did not bias measurement.

Item response theory

IRT was performed separately on each factor to evaluate the severity (difficulty), discrimination ability, and information (precision) of each item separately. Given the lack of substantial measurement invariance, we used the complete sample for these analyses (*n* = 548). Figure 2 depicts the item characteristic curves (ICC, for the severity and discrimination ability of the items) and the information curves (IFC, for the level of precision provided by the items) for each item within its factor. The severity and discrimination parameters are presented in Table 5 and the item and test information for each factor is shown in Supplementary Table 6.

Mind and body sleep problems

In terms of MBSP, "wanting to sleep better" was the least severe item and it was more likely to be reported by people with less than average MBSP scores. The most severe item was "felt too unsafe to sleep," which was more likely to be reported by individuals with above average MBSP scores. The MBSP items best able to discriminate between individuals with different severity of MBSP were "uncontrollable/racing thoughts when I tried to sleep" and "negative emotions when I tried to sleep." Thus, reporting or not reporting one of these two problems made the largest difference in the expected quality of sleep. These two items were also the most informative (precise) for individuals with average sleep quality ± 1 standard deviation (*SD*). For individuals with higher levels of MBSP, "felt too unsafe to sleep" and

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

Table 4. EFA and CFA loadings per sample—final model

		Sample					
		IP-a(EFA)		IP-b(CFA)		OL(CFA)	
Item	Description	MBSP	SRSP	MBSP	SRSP	MBSP	SRSP
I06	Uncontrollable/racing thoughts when I tried to sleep	**0.9**	0.0	**1.0**		**1.0**	
I07	Negative emotions when I tried to sleep	**0.9**	–0.1	**1.1**		**1.0**	
I02	Worried about my sleeping	**0.8**	–0.1	**0.9**		**0.9**	
I01	Wanted to sleep better	**0.9**	–0.1	**0.9**		**1.1**	
I05	Difficulty falling asleep	**0.8**	0.0	**0.9**		**0.9**	
I14	Panic attacks in the night	**0.6**	0.1	**0.8**		**0.7**	
I03	Woken up tired most mornings	**0.7**	0.1	**0.8**		**1.0**	
I10	Waking up lots in the night	**0.6**	0.0	**0.7**		**0.8**	
I19	Woken up feeling confused or disoriented	**0.7**	0.1	**0.8**		**0.7**	
I18	Felt restless in my sleep	**0.7**	0.2	**0.8**		**0.9**	
I04	Felt too unsafe to sleep	**0.6**	0.0	**0.7**		**1.0**	
I23	Too tired to think clearly or to do things during the day	**0.6**	0.3	**0.8**		**1.0**	
I16	Dreams which disturbed sleep	**0.7**	0.0	**0.7**		**0.6**	
I17	Aches and pains that stopped me from sleeping	**0.6**	0.1	**0.7**		**0.8**	
I12	Woken up in the night and used street drugs	0.0	**0.9**		**1.0**		**1.0**
I09	Street drugs to help me sleep	–0.1	**0.9**		**1.0**		**0.7**
I11	Woken up in the night and drunk alcohol	–0.1	**0.9**		**0.8**		**0.8**
I08	Drunk alcohol to help me sleep	0.1	**0.7**		**0.9**		**1.0**
I20	Woken up with a hangover or drunk	0.2	**0.6**		**0.9**		**1.1**
I22	Needed alcohol or drugs to get out of bed	0.1	**0.7**		**0.9**		**1.0**
I15	Vomited in my sleep*	0.2	**0.6**		**0.6**		-
I13	Woken up in the night and smoked tobacco	0.2	**0.5**		**0.6**		**0.7**
I21	Woken up withdrawing	0.2	**0.6**		**0.7**		**1.1**

All loadings marked with bold were significant. EFA loadings according to OBLIMIN rotation. In the OL sample, item I15 "vomited in my sleep" was omitted because there were no positive responses.

"experiencing panic attacks in the night" were the most informative items. For individuals with lower levels of MBSP, "wanting to sleep better" and "woken up tired most mornings" were the most informative.

The IFCs indicated that the MBSP items could measure problems precisely across the MBSP continuum; indeed, both the ICCs and the IFCs were spread across the MBSP latent continuum (Figure 2).

Substance-related sleep problems

Turning to SRSP, "vomited in my sleep" was the most severe item and "woken up in the night and smoked tobacco" was the least severe. "Woken up in the night and drunk alcohol," "woken up in the night and used street drugs," and "needed alcohol or drugs to get out of bed" were the most discriminative, although most items performed similarly well. For individuals with up to average SRSP, "needing alcohol or drugs to get out of bed" was the most informative item. For higher levels of SRSP, most IFCs peaked at about 1 *SD* above average. That is, the precision of the items was maximized for individuals with SRSP scores close to 1 *SD* above average. "Vomiting in my sleep" had low precision for individuals with low to up to 1 *SD* above average SRSP levels. However, for high SRSP levels, "vomited in my sleep" was the most informative item; indicating that this question was particularly useful for identifying individuals with severe substance-related problems.

Whilst the ICCs of MBSP were spread over the latent continuum (± 2 *SD* from the mean), the characteristic curves of the SRSP items were located at above average sleep problems. This indicated that SRSP items pertained largely to very poor quality sleep.

Reliability and validity

Reliability

Reliability tests on the final 23-item sleep PROM showed that Cronbach's α coefficient was still 0.88 for the complete sample, with no improvement in the reliability index gained by omitting items. The item–total correlations varied between $r = 0.3$ and $r = 0.6$. For MBSP, α was 0.86 (item–item correlations: 0.2–0.6; item–total correlations: 0.4–0.6) and no problematic items were found. For SRSP, which comprised a smaller number of items, α was 0.79 (item–item correlations: 0.14–0.7; item–total correlations: 0.3–0.6) and no problematic items were found. Thus, internal consistency was granted for the final measure in the complete sample.

With respect to the stability of the final measure, all items had fair-to-perfect agreement as reported previously (Table 2; Cohen's κ varied between 0.4 and 1) [32]. With respect to the factor and total scores, the intraclass correlation coefficient indicated almost perfect agreement between the two time points in all cases (MBSP: 0.96, 95% CI: [0.92, 0.98]; SRSP: 0.97, 95% CI: [0.95, 0.99], TS: 0.96, 95% CI: [0.92, 0.98]). We therefore concluded that the measure was reliable, both according to classical test theory (internal consistency, stability) and IRT.

Validity

In terms of validity, the new measure correlated moderately to strongly with the PSQI and SURE scores demonstrating convergent validity. In contrast, there were only low or nonsignificant correlations with actigraphy scores indicating discriminant validity (Table 6).

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

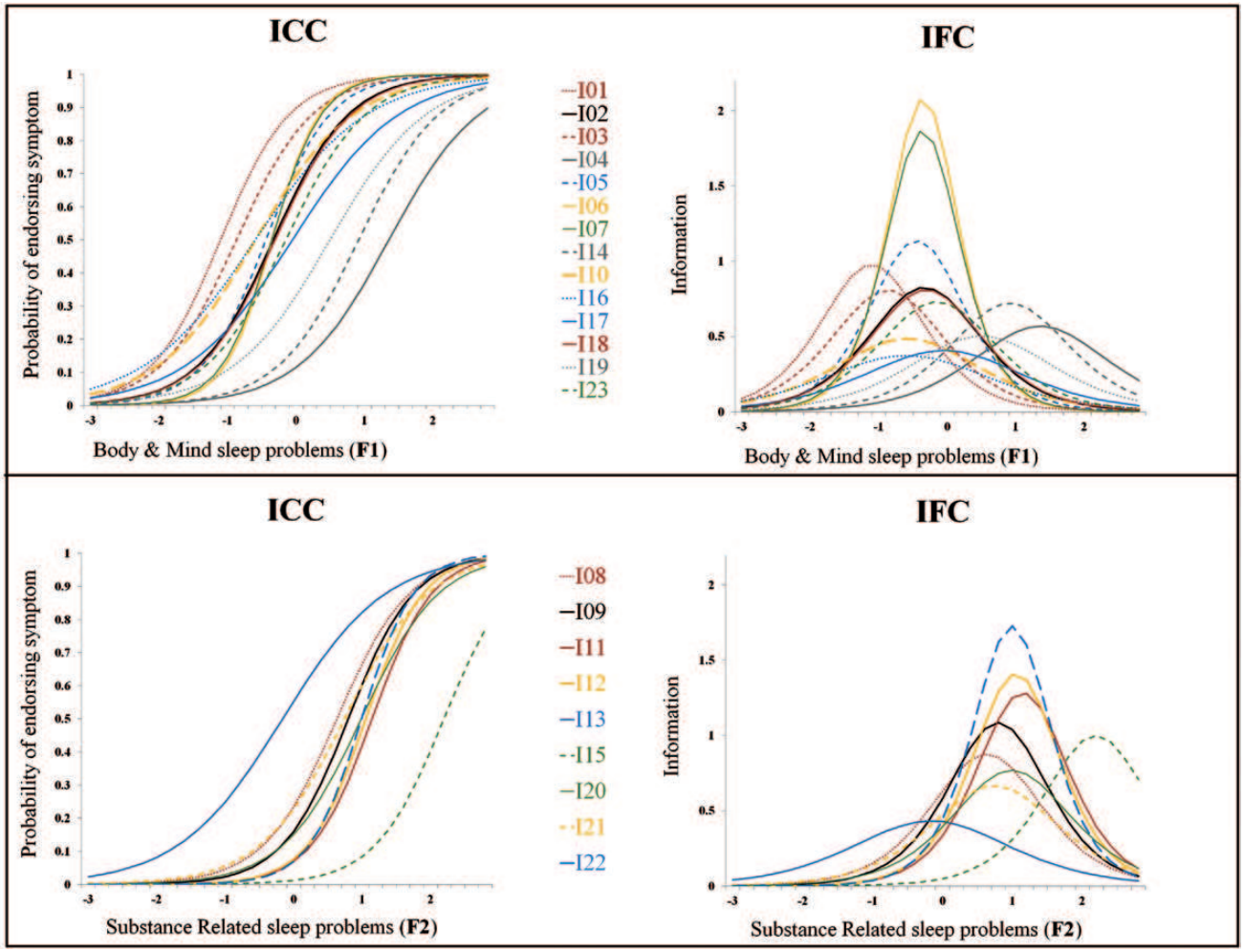

Figure 2. Item characteristic curves (ICC) and item information curves (IFC) for each factor.

Table 5. IRT parameters ($p < 0.05$ in all cases)

	Item	Discrimination	Difficulty
MBSP	I01	1.98	−1.10
	I02	1.82	−0.33
	I03	1.80	−0.87
	I04	1.51	1.36
	I05	2.14	−0.44
	I06	2.89	−0.35
	I07	2.74	−0.36
	I14	1.70	0.91
	I10	1.39	−0.59
	I16	1.22	−0.59
	I17	1.28	−0.05
	I18	1.80	−0.31
	I19	1.42	0.50
	I23	1.71	−0.15
SRSP	I08	1.87	0.63
	I09	2.09	0.80
	I11	2.27	1.14
	I12	2.38	1.05
	I13	1.32	−0.16
	I15	2.00	2.20
	I20	1.75	0.99
	I21	1.63	0.74
	I22	2.63	0.99

The new measure was also able to discriminate between subgroups of substance users (Table 7). For example, people who reported insomnia had significantly higher factor scores than those who did not report insomnia (both factors); people who had been homeless in the last month scored significantly higher than those who had not been homeless in the last month (both factors); and people who had been in paid work in the last month scored significantly lower than those who had not been in paid work in the last month (both factors).

Analyses presented in Table 7 also provide further evidence that the two factors are distinct in terms of content. Thus, people who reported mental health problems had significantly higher MBSP scores than those who did not report mental health problems, but the two groups did not have significantly different SRSP scores. Meanwhile, gender and education were not related to MBSP, but were related to SRSP.

Discussion

Sleep is an important, but frequently overlooked, issue for people addicted to alcohol and other drugs. Although there are many validated sleep measures, none has been developed specifically for people experiencing problems with substances and, consequently, there has been no reliable way of assessing the self-reported sleep problems of this population. We completed

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

extensive developmental work, comprising qualitative and quantitative methods, with significant input from people using drugs and alcohol in order to develop our new self-reported outcome measure of sleep problems. Members of our service user PAG confirmed face and content validity and determined the measure's name: the SUSS. Statistical analyses established a two-factor structure, measurement invariance, reliability, and validity.

SUSS is shown in Table 8. It comprises 23 items and two factors: factor 1, "Mind and Body Sleep Problems" and factor 2, "Substance Related Sleep Problems." Each factor is internally coherent. Items score 0 or 1, so total scores range from 0 to 23 (where lower scores denote better sleep and higher scores denote worse sleep). SUSS is worded so that it is accessible to people with limited literacy: it has a Flesch readability score (based on number of words per sentence and number of syllables per word) of 86.8 and a Lexile Measure (based on word frequency and sentence length) of 400L–500L. These scores denote easy to read conversational English.

SUSS is quick and easy to complete (as little as 2 min). However, individuals frequently took longer to complete it because they chose to pause and reflect on individual items and to discuss their responses with the researcher. As has been argued previously [25, 29], a well-designed PROM will not simply generate numeric scores. It will also prompt people to reflect on and volunteer potentially important information about their lives and circumstances that might be helpful in a therapeutic context. Since only 1/30 people reported that the initial 30-item version of SUSS was too long (whereas 2/30 reported it was too short), 20/30 reported that they enjoyed completing it, and 28/30 identified no irrelevant questions (Item development, stage 3), we do not feel that a short-form version is an immediate priority for clinical practice. Nonetheless, we recognize that brevity is important when scales are used in studies alongside other assessments. IRT is not designed as a method to reduce the number of items in a measure, but our IRT parameters fully describe the functionality of each item and may thus facilitate item selection in the development of any future short form.

Factor 1 ("Mind and Body Sleep Problems") comprises 14 items that all relate to cognitive and behavioral difficulties and concerns about sleep. These include problems going to sleep ("uncontrollable/racing thoughts when trying to sleep," "negative emotions when trying to sleep," "difficulty falling asleep," "feeling too unsafe to sleep," "aches and pains that prevented

Table 6. Pearson correlation coefficients for the final 23-item sleep PROM

	MBSP scores			SRSP scores			Total score		
	r	*P*	*N*	*r*	*P*	*N*	*r*	*P*	*N*
Age	**–0.1**	0.001	547	**–0.1**	0.005	547	**–0.2**	<0.001	547
PSQI score	**0.7**	<0.001	100	**0.3**	0.001	100	**0.7**	<0.001	100
SURE score	**–0.5**	<0.001	100	**–0.6**	<0.001	100	**–0.6**	<0.001	100
Actigraphy: Sleep efficiency	**–0.4**	0.044	22	–0.3	0.197	22	**–0.5**	0.022	22
Actigraphy: Sleep latency	0.2	0.283	22	0.2	0.263	22	0.3	0.145	22
Actigraphy: Total sleep time	–0.4	0.039	22	–0.1	0.569	22	–0.4	0.062	22
Actigraphy: Wake bouts	0.1	0.768	22	0.0	0.913	22	0.0	0.874	22

Table 7. Score differences in relation to demographic and clinical characteristics for the final 23-item sleep PROM

	MBSP		SRSP		SP	
Factor	95% CI	Comparison	95% CI	Comparison	95% CI	Comparison
Gender (Male–Female)	(–0.7, 0.6)	$t = –0.1$, df–547, $p = 0.922$	(0.5, 1.3)	**W = 42915.5, $p < 0.001$**	(0, 1.8)	$t = 1.9$, df–547, $p = 0.057$
Ethnicity (White–Other)	(–0.9, 0.8)	$t = 0$, df–547, $p = 0.962$	(–0.9, 0.1)	W = 19861, $p = 0.086$	(–1.6, 0.8)	$t = –0.7$, df–547, $p = 0.492$
Education beyond 16 years (N–Y)	(–0.2, 1.1)	$t = 1.4$, df–545, $p = 0.148$	(0.8, 1.6)	W = 46800, $p < 0.001$	(0.8, 2.6)	$t = 3.7$, df–545, $p < 0.001$
Substance use in the last 6 months (Y–N)	(3.5, 1.9)	**$t = 6.8$, df–544, $p < 0.001$**	(2.8, 1.9)	**W = 7583, $p \leq 0.001$**	(6.1, 4)	**$t = 9.5$, df–544, $p < 0.001$**
Homeless in last month (Y–N)	(0.7, 2.6)	**$t = 3.5$, df–547, $p < 0.001$**	(0.8, 1.9)	**W = 23805.5, $p < 0.001$**	(1.7, 4.3)	**$t = 4.7$, df–547, $p < 0.001$**
Had paid work in last month (Y–N)	(–1.9, –0.2)	**$t = –2.4$, df–547, $p = 0.017$**	(–1.5, –0.4)	**W = 16065, $p < 0.001$**	(–3.2, –0.8)	**$t = –3.3$, df–547, $p = 0.001$**
Has a diagnosed physical health problem (Y–N)	(–0.4, 2.7)	$t = 1.5$, df–98, $p = 0.133$	(–1, 0.9)	W = 1204, $p = 0.816$	(–1, 3.1)	$t = 1.1$, df–98, $p = 0.293$
Has a diagnosed mental health problem (Y–N)	(1.4, 4.3)	**$t = 4$, df–98, $p < 0.001$**	(–0.5, 1.5)	W = 1343, $p = 0.235$	(1.4, 5.4)	**$t = 3.3$, df–98, $p = 0.001$**
Ever diagnosed insomnia (Y/N)	(1.8, 3.4)	**$t = 6.4$, df–522, $p < 0.001$**	(–0.1, 0.9)	**W = 24339, $p = 0.066$**	(1.9, 4.1)	**$t = 5.4$, df–522, $p < 0.001$**

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

sleeping"); problems staying asleep ("panic attacks in the night," "waking up lots in the night," "being restless during sleep," "dreams that disturbed sleep"); and problems associated with fatigue on wakening and during the day ("waking up tired most mornings," "waking up feeling confused and disoriented," "being too tired to think clearly or do things during the day," "worrying about sleep," "wanting to sleep better"). None of the 14 items explicitly mentions drugs or alcohol and all items could be reported by individuals who do not consume substances. Nevertheless, the factor captures aspects of sleep (particularly, "racing thoughts," "feeling unsafe," "panic attacks," "disturbing dreams," and "waking up confused and disoriented") that are germane to substance users [2, 3] and tend not to feature in other more generic sleep scales.

Items in Factor 2 ("Substance-Related Sleep Problems") also relate to difficulties when trying to go to sleep, difficulties staying asleep, and difficulties on wakening. However, in contrast to factor 1, all factor 2 items (n = 9) refer directly or indirectly to substances (drugs, alcohol, or tobacco) and none would ever be reported by anyone who did not consume substances. Thus, five items explicitly mention drugs or alcohol ("waking up in the night and using street drugs," "using street drugs to help sleep," "waking up in the night and drinking alcohol," "drinking alcohol to help sleep," "needing alcohol or drugs to get out of bed") and one item explicitly mentions tobacco ("waking up in the night and smoking tobacco"). The remaining three items ("woken up with a hangover," "vomited in my sleep," "woken up withdrawing") all clearly refer to the consequences of drug or alcohol consumption and intoxication.

People experiencing problems with drugs and alcohol often report complex physical, psychological, and social problems that can impede sleep [3, 27]. Within the general population, biopsychosocial problems also undermine sleep [46]. We might therefore expect that people who report problematic substance use will experience at least some sleep difficulties independent of their alcohol and other drug consumption such that sleep problems will not simply disappear if they become abstinent. This assumption is supported by the two-factor structure of SUSS which distinguishes more generalized MBSP from sleep problems related specifically to substance use. Reflecting this, we suggest that any interventions or strategies designed to improve the sleep of people who use drugs or alcohol will need to address both their substance use and any wider biopsychosocial problems that patient and clients report; addressing substance use alone will likely generate only partial success [47].

Invariance testing revealed that participants who reported the same level of MBSP were less likely to report "difficulty falling asleep" and "negative emotions when they tried to sleep" but were more likely to report "aches and pains that stopped them from sleeping" as they aged. Meanwhile, participants with the same level of substance-related sleep problems were more likely to report "drinking alcohol to help them sleep" and "waking up in the night to drink alcohol" as they aged. These findings suggest that older people reporting problems with substances may experience sleep problems differently from younger people. This may limit the ability of SUSS to measure and compare sleep problems across age groups. Nonetheless, the magnitude of the effects was low and changes to sleep patterns are part of the normal aging process. Studies have consistently shown a decrease in both sleep length and slow-wave sleep and an increase in sleep fragmentation in older adults [48, 49], as well as adaptation in perceptions of sleep and disturbances as we age [50]. Some basic differences in perceptions of sleep quality

Table 8. Substance Use Sleep Scale (SUSS)

Mind and body sleep problems—*thinking about the last week*	No	Yes
1. I have worried about my sleeping	☐	☐
2. I have wanted to sleep better	☐	☐
3. I have had difficulty falling asleep	☐	☐
4. I have felt too unsafe to sleep	☐	☐
5. I have had uncontrollable/racing thoughts when I tried to sleep	☐	☐
6. I have had negative emotions (such as anger, guilt, or anxiety) when I tried to sleep	☐	☐
7. I have had aches and pains that stopped me from sleeping	☐	☐
8. I have been waking up lots in the night	☐	☐
9. I have had panic attacks in the night	☐	☐
10. I have had dreams which have disturbed my sleep	☐	☐
11. I have felt restless in my sleep (e.g. jumpy, twitchy, or itchy legs)	☐	☐
12. I have woken up feeling confused or disoriented	☐	☐
13. I have woken up tired most mornings	☐	☐
14. I have been too tired to think clearly or to do things during the day	☐	☐
Substance-related sleep problems—*thinking about the last week*	**No**	**Yes**
15. I have drunk alcohol to help me sleep	☐	☐
16. I have taken street drugs to help me sleep	☐	☐
17. I have woken up in the night and drunk alcohol	☐	☐
18. I have woken up in the night and used street drugs	☐	☐
19. I have woken up in the night and smoked tobacco	☐	☐
20. I have vomited in my sleep	☐	☐
21. I have woken up with a hangover or drunk	☐	☐
22. I have woken up withdrawing	☐	☐
23. I have needed alcohol or drugs to get out of bed	☐	☐

Items score 0 (No) or 1 (Yes), so total scores range from 0 to 23 (where lower scores denote better sleep and higher scores denote worse sleep).

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

between age groups are therefore to be expected and do not notably undermine the utility of SUSS.

The IRT models indicated that items relating to MBSP captured high, medium, and low levels of sleep difficulty, whereas items relating to SRSP mainly captured very poor sleep quality. Thus, people who have low SRSP scores (i.e. people who report that substance use is negatively affecting their sleep) are likely to have especially poor sleep. Addressing their substance use may significantly improve (although not necessarily solve) their sleep problems. That the only item relating to tobacco "woken up and smoked tobacco" captured sleep quality at the higher end of substance-related sleep trait suggests that alcohol and other drug use are more closely associated with poor sleep than tobacco. Consequently, addressing alcohol and other drugs seems more likely to improve sleep than addressing smoking, although reduced use of tobacco should still improve sleep quality [51].

SUSS, the PSQI, and SURE are all self-report measures assessing self-perceived health status, whereas actigraphy scores are a more objective measure of sleep based on movement. Given that subjective and actigraphic measurements of sleep are known to correlate poorly [52–54], our finding that SUSS correlated moderately to strongly with the PSQI and SURE scores but not with the actigraphy scores are as anticipated. Similarly, we expected and found that SUSS correlated positively with two characteristics likely to be associated with poor sleep (insomnia and homelessness) and negatively with one characteristic likely to be associated with good sleep (being in paid work). Meanwhile, gender and education were associated with SRSP but not MBSP, whereas having a diagnosed mental health problem was related to MBSP but not SRSP. These latter findings merit further consideration.

Although there is evidence of sex differences in sleep difficulties [55], analysis of large-scale survey data has suggested that a major reason why women self-report poorer sleep than men relates to women's disadvantaged socioeconomic status [56]. Supporting this, Sekine et al. have found that gender differences in sleep could be entirely explained by gender differences in work characteristics, domestic roles, and family work conflicts [57]. The lack of association between gender and MBSP in our analyses may therefore relate to the specific socioeconomic patterns and dynamics of alcohol and other drug users. For example, poor education, low income, and unemployment tend to be high across both sexes, with some evidence that women may be better than men at managing the finances they do have once substance use stops [3, 58]. In terms of differences between mental health and the two sleep factors, there is a well-known association between poor mental health and the general sleep issues comprising factor 1 [59, 60]. There is, however, no current evidence of an association between mental health and the items comprising SRSP (factor 2).

Limitations and strengths

Self-report measures of sleep have limitations despite their widespread use [5]. Awareness is reduced during sleep and this limits the validity of sleep self-reports when assessing objective variables and phenomena such as snorting, apnea, or leg jerks [5]. Sleep can also vary considerably from day to day, meaning that multiple measures are often needed to derive stable measures [5]. Further validity testing of SUSS, based on larger and more diverse samples that would establish factor and total score norms and enable other group comparisons, is now needed. This might include people who consume alcohol or other drugs but do not consider that they have ever had a problem with substances; people from other patient populations who report sleep problems; and people from other countries and cultures where language, practices, and expectations relating to both sleep and substance use may vary. In addition, criterion and predictive validity testing would be desirable.

In terms of its strengths, SUSS is the first sleep measure designed specifically for people experiencing problems with alcohol and other drugs. Moreover, people who used substances played a central role in establishing its content, and confirmed that it had good face validity. The measure was designed using a careful and considered blending of qualitative methods (with their focus on subjective meaning and understanding) and more objective quantitative techniques. This capitalized on the strengths of each to ensure a robust development and early validation process. Our study participants were diverse in terms of age, ethnicity, drugs used, and geographical location, so providing reassurances in terms of inclusion and diversity. That we successfully collected data OL suggests that a computer adaptive version can be designed. Lastly, SUSS seems likely to have a number of potentially important functions. It might be used by people who use, or have used, substances to monitor and reflect on their own sleep; by treatment providers to encourage and enable people who use substances to think about their sleep and its diverse correlates, and to practice, evaluate, and deploy strategies for improving sleep; and by researchers and others as an outcome measure when designing and implementing sleep interventions for this population.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Acknowledgments

The authors thank all individuals who participated in the research and all services and staff for providing access to their clients. We also acknowledge Zoe Gotts for her work identifying published sleep measures in stage 1 of the study whilst she was employed on the project "Sleep during recovery from drug and alcohol dependence: developing an interdisciplinary approach and exploring the feasibility of an intervention" (funded by Wellcome Trust [ref: 105624] through the Centre for Chronic Diseases and Disorders [C2D2] at the University of York). The qualitative data used for the secondary analysis were collected with support from a British Academy/Leverhulme Small Research Grant Scheme SRG 2013-14 (Reference: SG132091). Sleep during recovery from drug and alcohol dependence: a sociological study of embodied change. In addition, we are grateful to members of our two project advisory groups for their support and guidance, Action on Addiction for funding the research, and Nick Barton for his continued support and enthusiasm for all of our work on sleep and drug and alcohol dependence.

Funding

This research was undertaken with financial support from Action on Addiction. J.S. is supported by, and J.N. and S.V. are partly funded by, the National Institute for Health Research

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

(NIHR) Biomedical Research Centre for Mental Health at South London, Maudsley NHS Foundation Trust, and King's College London.

Notes

Conflict of interest statement. J.S. is a researcher and clinician who has worked with a range of types of treatment and rehabilitation service-providers. He has also worked with a range of governmental and nongovernmental organizations, and with pharmaceutical companies to seek to identify new or improved treatments from whom he and his employer (King's College London) have received honoraria, travel costs, and/or consultancy payments. This includes work with, during past 3 years, Martindale, Indivior, MundiPharma, Braeburn/Camurus, and trial medication supply from iGen and from Camurus. His employer (King's College London) has registered intellectual property on a novel buccal naloxone formulation and he has also been named in a patent registration by a Pharma company regarding a concentrated nasal naloxone spray. For a fuller account, see JS's web-page at http://www.kcl.ac.uk/ioppn/depts/addictions/people/hod.aspx. The authors have no other conflicts of interest. The views expressed are those of the authors and not necessarily those of Action on Addiction, the NHS, the NIHR, the Department of Health, or Public Health England.

References

1. Arnedt JT, *et al*. Sleep and substance use disorders. In: Morin CM, Espie CA, eds. *The Oxford Handbook of Sleep and Sleep Disorders*. Oxford: Oxford University Press; 2012.
2. Nettleton S, *et al*. Techniques and transitions: a sociological analysis of sleeping practices amongst recovering heroin users. *Soc Sci Med*. 2011; **72**(8): 1367–1373.
3. Neale J, *et al*. *The Everyday Lives of Recovering Heroin Users*. London: Royal Society of Arts; 2012.
4. Neale J, *et al*. "You're all going to hate the word 'recovery' by the end of this": service users' views of measuring addiction recovery. *Drugs-educ prev polic*. 2015; **22**(1): 26–34.
5. Buysse DJ, *et al*. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010; **33**(6): 781–792.
6. Foster RG,*et al*. *Sleep: A Very Short Introduction*. Oxford: Oxford University Press; 2012.
7. Breslau N, *et al*. Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. *Biol Psychiatry*. 1996; **39**(6): 411–418.
8. Spiegel K, *et al*. Sleep loss: a novel risk factor for insulin resistance and type 2 diabetes. *J Appl Physiol (1985)*. 2005; **99**(5): 2008–2019.
9. Cappuccio FP, *et al*. Gender-specific associations of short sleep duration with prevalent and incident hypertension: the Whitehall II Study. *Hypertension*. 2007; **50**(4): 693–700.
10. Cappuccio FP, *et al*. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep*. 2008; **31**(5): 619–626.
11. Hall MH, *et al*. Self-reported sleep duration is associated with the metabolic syndrome in midlife adults. *Sleep*. 2008; **31**(5): 635–643.
12. Lyznicki JM, *et al*. Sleepiness, driving, and motor vehicle crashes. *JAMA*. 1998; **279**(23): 1908–1913.
13. Walker MP, *et al*. Sleep-dependent learning and memory consolidation. *Neuron*. 2004; **44**(1): 121–133.
14. Gangwisch JE, *et al*. Sleep duration associated with mortality in elderly, but not middle-aged, adults in a large US sample. *Sleep*. 2008; **31**(8): 1087–1096.
15. Gallicchio L, *et al*. Sleep duration and mortality: a systematic review and meta-analysis. *J Sleep Res*. 2009; **18**(2): 148–158.
16. Oyefeso A, *et al*. Subjective sleep-wake parameters in treatment-seeking opiate addicts. *Drug Alcohol Depend*. 1997; **48**(1): 9–16.
17. Stein MD, *et al*. Sleep disturbances among methadone maintained patients. *J Subst Abuse Treat*. 2004; **26**(3): 175–180.
18. Morgan PT, *et al*. Sleep, sleep-dependent procedural learning and vigilance in chronic cocaine users: evidence for occult insomnia. *Drug Alcohol Depend*. 2006; **82**(3): 238–249.
19. Burke CK, *et al*. Sleep problems reported by patients entering opioid agonist treatment. *J Subst Abuse Treat*. 2008; **35**(3): 328–333.
20. Escobar-Córdoba F, *et al*. Complaints of insomnia in hospitalized alcoholics. *Rev Bras Psiquiatr*. 2009; **31**(3): 261–264.
21. Mahfoud Y, *et al*. Sleep disorders in substance abusers: how common are they? *Psychiatry (Edgmont)*. 2009; **6**(9): 38–42.
22. Arnedt JT, *et al*. Cognitive-behavioral therapy for insomnia in alcohol dependent patients: a randomized controlled pilot trial. *Behav Res Ther*. 2011; **49**(4): 227–233.
23. Dawson J. Measuring health status. In: Neale J, ed. *Research Methods for Health and Social Care*. London: Palgrave; 2009: 181–194.
24. Patrick D, *et al*. Patient-reported outcomes. In: Higgens J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons; 2008: 531–545.
25. Neale J, *et al*. Philosophical ruminations on measurement: methodological orientations of patient reported outcome measures (PROMS). *J Ment Health*. 2015; **24**(3): 123–125.
26. Meadows R, Nettleton S, Neale J. Sleep waves and recovery from drug and alcohol dependence: Towards a rhythm analysis of sleep in residential treatment. *Soc Sci Med*. 2017; **184**: 124–133.
27. Nettleton S, *et al*. Disturbing sleep and sleepfulness during recovery from substance dependence in residential rehabilitation settings. *Sociol Health Illn*. 2017; **39**(5): 784–798.
28. Buysse DJ, *et al*. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989; **28**(2): 193–213.
29. Neale J, *et al*. Development and Validation of 'SURE': a patient reported outcome measure (PROM) for recovery from drug and alcohol dependence. *Drug Alcohol Depend*. 2016; **165**: 159–167.
30. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; **16**(3): 297–334.
31. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; **20**(1): 37–46.
32. Landis JR, *et al*. The measurement of observer agreement for categorical data. *Biometrics*. 1977; **33**(1): 159–174.
33. Shrout PE, *et al*. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; **86**(2): 420–428.
34. Muthén BO. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984; **49**(1): 115–132.
35. Jöreskog KG. Statistical analysis of sets of congeneric tests. *Psychometrika*. 1971; **36**(2): 109–133.
36. Byrne BM, *et al*. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol Bull*. 1989; **105**(3): 456–466.
37. Millsap RE. *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge; 2011.

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018

38. Jöreskog K, *et al*. Estimation of a model with a multiple indicators and multiple causes of a single latent variable. *J Am Stat Assoc*. 1975; **70**(351): 631–639.
39. Muthén BO. Latent variable modelling in heterogeneous populations. *Psychometrika*. 1989; **54**(4): 557–585.
40. Baker F. *The Basics of Item Response Theory*. University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation; 2001.
41. Hoelter JW. The analysis of covariance structures: goodness of fit indices. *Sociol Methods Res* 1983; **11**: 325–344.
42. Browne MW and Cudeck, R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Newbury Park: Sage; 1993.
43. Bentler PM and Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull* 1980; **88**: 588–606.
44. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990; **107**: 238–246
45. Broderick JE, Junghaenel DU, Schneider S, Pilosi JJ, Stone AA. Pittsburgh and Epworth sleep scale items: accuracy of ratings across different reporting periods. *Behav Sleep Med*. 2013; **11**(3): 173–188.
46. Dregan A, *et al*. Potential pathways from biopsychosocial risk factors to sleep loss due to worry: a population-based investigation. *J Public Ment Health*. 2013; **12**(1): 43–50.
47. Neale J, *et al*. Substance misuse, sleep and intervention design: insights from qualitative data. *J Ment Health*. In press.
48. Ohayon MM, *et al*. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*. 2004; **27**(7): 1255–1273.
49. Schwarz JFA, *et al*. Age affects sleep microstructure more than sleep macrostructure. *J Sleep Res*. 2017; **26**(3): 277–287.
50. Buysse DJ, *et al*. Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh Sleep Quality Index (PSQI). *Sleep*. 1991; **14**(4): 331–338.
51. Jaehne A, *et al*. How smoking affects sleep: a polysomnographical analysis. *Sleep Med*. 2012; **13**(10): 1286–1292.
52. Lockley SW, *et al*. Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. *J Sleep Res*. 1999; **8**(3): 175–183.
53. Kushida CA, *et al*. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001; **2**(5): 389–396.
54. Unruh ML, *et al*. Subjective and objective sleep quality and aging in the sleep heart health study. *J Am Geriatr Soc*. 2008; **56**(7): 1218–1227.
55. Zhang B, *et al*. Sex differences in insomnia: a meta-analysis. *Sleep*. 2006; **29**(1): 85–93.
56. Arber S, *et al*. Gender and socio-economic patterning of self-reported sleep problems in Britain. *Soc Sci Med*. 2009; **68**(2): 281–289.
57. Sekine M, *et al*. Work and family characteristics as determinants of socioeconomic and sex inequalities in sleep: The Japanese Civil Servants Study. *Sleep*. 2006; **29**(2): 206–216.
58. Neale J, *et al*. Gender sameness and difference in recovery from heroin dependence: a qualitative exploration. *Int J Drug Policy*. 2014; **25**(1): 3–12.
59. Augner C. Associations of subjective sleep quality with depression score, anxiety, physical symptoms and sleep onset latency in students. *Cent Eur J Public Health*. 2011; **19**(2): 115–117.
60. Dolsen MR, *et al*. Life-time history of insomnia and hypersomnia symptoms as correlates of alcohol, cocaine and heroin use and relapse among adults seeking substance use treatment in the United States from 1991 to 1994. *Addiction*. 2017; **112**(6): 1104–1111.

Downloaded from https://academic.oup.com/sleep/advance-article-abstract/doi/10.1093/sleep/zsy013/4796928
by University of York user
on 14 March 2018