BMJ Open

Prescription medications for sleep disturbances among midlife women during 2 years of follow-up: a SWAN retrospective cohort study

Daniel H Solomon ,[1,2] Kristine Ruppert,[3] Laurel A Habel,[4] Joel S Finkelstein,[5] Pam Lian,[3] Hadine Joffe,[6] Howard M Kravitz[7]

To cite: Solomon DH, Ruppert K, Habel LA, *et al*. Prescription medications for sleep disturbances among midlife women during 2 years of follow-up: a SWAN retrospective cohort study. *BMJ Open* 2021;**11**:e045074. doi:10.1136/bmjopen-2020-045074

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-045074).

Received 22 September 2020
Revised 23 February 2021
Accepted 24 February 2021

For numbered affiliations see end of article.

Correspondence to
Dr Daniel H Solomon;
dsolomon@bwh.harvard.edu

ABSTRACT

Objective To examine the effects of prescription sleep medications on patient-reported sleep disturbances.

Design Retrospective cohort.

Setting Longitudinal cohort of community-dwelling women in the USA.

Participants Racially and ethnically diverse middle-aged women who reported a sleep disturbance.

Interventions New users of prescription sleep medications propensity score matched to women not starting sleep medications.

Main outcomes and measures Self-reported sleep disturbance during the previous 2 weeks—difficulty initiating sleep, waking frequently and early morning awakening—using a 5-point Likert scale, ranging from no difficulty on any night (rating 1) to difficulty on 5 or more nights a week (rating 5). Sleep disturbances were compared at 1 year (primary outcome) and 2 years of follow-up.

Results 238 women who started sleep medications were matched with 447 non-users. Participants had a mean age of 49.5 years and approximately half were white. At baseline, sleep disturbance ratings were similar: medication users had a mean score for difficulty initiating sleep of 2.7 (95% CI 2.5 to 2.9), waking frequently 3.8 (95% CI 3.6 to 3.9) and early morning awakening 2.8 (95% CI 2.6 to 3.0); non-users ratings were 2.6 (95% CI 2.5 to 2.7), 3.7 (95% CI 3.6 to 3.9) and 2.7 (95% CI 2.6 to 2.8), respectively. After 1 year, ratings for medication users were 2.6 (95% CI 2.4 to 2.8) for initiating sleep, 3.6 (95% CI 3.4 to 3.8) for waking frequently and 2.8 (95% CI 2.6 to 3.0) for early morning awakening; for non-users, the mean ratings were 2.3 (95% CI 2.2 to 2.5), 3.5 (95% CI 3.3 to 3.6) and 2.5 (95% CI 2.3 to 2.6), respectively. None of the 1 year changes were statistically significant nor were they different between medication users and non-users. Two-year follow-up results were consistent, without statistically significant reductions in sleep disturbance in medication users compared with non-users.

Conclusions These analyses suggest that women who initiated sleep medications rated their sleep disturbances similar after 1 and 2 years. The effectiveness of long-term sleep medication use should be re-examined.

Strengths and limitations of this study

- ► Little is known about the long-term benefits of medications used for sleep in typical practice.
- ► We compared reductions in sleep difficulties across a large cohort of women reporting sleep difficulties who did and did not start prescription medications used for sleep.
- ► Some of these medications may not have been prescribed for sleep difficulties and some medications were likely used intermittently.

INTRODUCTION

Sleep disturbances are common, and an estimated 9 million adults in the USA report prescription medication use for this indication.[1] The frequency of sleep medication use has increased since the 1990s and first decade of the 2000s.[2 3] Sleep disorders are associated with many important chronic conditions, including diabetes, hypertension, pain and depression.[4] Due to the prevalence of sleep disturbances and their interplay with important comorbidities, many pharmacological treatment options have been developed for sleep.

Prescription sleep medications consist of benzodiazepines (BZDs), Z-drugs (selective BZD receptor agonists that include zolpidem, zaleplon and eszopiclone) and other agents mostly used off-label to promote sleep through a variety of other mechanisms. Randomised controlled trials (RCTs) demonstrate the short-term sleep benefits of many agents in these categories, with typical trials for these agents lasting only 12–24 weeks and often including fewer than 100 patients.[5 6] One 8-month study of zolpidem found improved polysomnographic sleep parameters and subject assessments on two nights in month 8.[7] While sleep medications are recommended for short courses,[8] sleep disturbances may be chronic and many patients use these agents for long periods, sometimes intermittently and other times

nightly.[9] Thus, data from typical practice would be useful for patients and clinicians if it included sleep medications used over several months in populations of patients with sleep disturbances; we found no such studies in the literature.

There has been increased interest in using non-randomised designs to test the benefits of drugs.[10] We assessed the potential benefits of sleep medications among a large and diverse cohort of midlife women not reporting prevalent sleep medication use at baseline who self-reported sleep disturbances during observation in a longitudinal cohort. Women who subsequently started sleep medications were matched on a propensity score with women who did not and followed for 1–2 years with annual assessment of sleep disturbances.

METHODS

Study design

The design of this study was based on the 'target trial emulation' concept as proposed by Hernán and Robins.[11] In this study paradigm, a target RCT is designed and then an observational study is constructed to emulate the target trial. We specified all relevant aspects of the target trial and the observational corollary as noted in online supplemental table 1. The observational study focused on new users of sleep medications, never previously reporting sleep medication use during the period of observation and primarily used an intention-to-treat design to most closely emulate the target trial. Furthermore, we described the study design using standardised illustrations as suggested by Schneeweiss *et al* (see online supplemental figure 1).[12]

Setting and participants

All potentially eligible women were drawn from the Study of Women's Health Across the Nation (SWAN). SWAN is an ongoing multicentre, multiethnic/multiracial longitudinal study examining the biological and psychosocial changes that occur during the menopausal transition. Between 1995 and 1997, a screening survey assessed the eligibility of women at each of seven participating sites; sampling used either community-based or population-based frames.[13] Major cohort entry criteria included: age 42–52 years; intact uterus and at least one ovary, not using sex steroid hormones or pregnant, breast feeding or lactating at enrolment or within the previous 3 months; at least one menstrual period in the 3 months prior to screening and self-identified as either white, African-American, Hispanic, Chinese or Japanese. Each site recruited at least 450 eligible women, including white women and a minority group sample, into the cohort in 1995–1997, resulting in an inception cohort of 3302 women.[14 15] For the current analyses, we used follow-up data through 2016.

Since we were interested in the long-term effects of prescription sleep medications on sleep disturbances, we required all women to have reported during SWAN follow-up a sleep disturbance on at least three nights per week during a 2-week interval. On almost all annual visits, women were asked to self-report on three aspects of sleep: difficulty initiating, frequent awakening and early morning awakening. If women reported any of these disturbances at least once, they were eligible for the study cohort. We also required women to have sleep data at the visit after first reporting a sleep disturbance; some visits did not include the brief sleep inventory and thus follow-up information would be missing. Finally, we excluded women who reported use of prescription sleep medications at the baseline visit in SWAN, to eliminate prevalent users of these drugs.

Patient and public involvement

There was no patient or public involvement in this research. Participants in SWAN receive updates on the conduct and results of the study. Data from SWAN are available for qualified researchers. All participants gave written informed consent to use their data for these analyses. The current analyses were funded by the US National Institutes of Health. All participants gave written informed consent after being educated about the nature of the study, potential risks and how their data may be used.

Exposures

Many different medications are used for sleep. We focused on several groups of medications: BZDs, selective BZD receptor agonists and other hypnotics. The full list of medications considered included the following BZDs: estazolam, flurazepam, lorazepam, temazepam and triazolam;, selective BZD receptor agonists: zaleplon, zolpidem and eszopiclone and agents with other mechanisms: doxepin (a tertiary amine tricyclic), mirtazapine (noradrenergic and specific serotonergic), ramelteon (selective melatonin receptor agonist) and trazodone (serotonin antagonist and reuptake inhibitor). The primary analyses grouped all sleep medications together. In secondary analyses, groups of medications were considered separately. Lorazepam users (n=65) and their matched non-users (n=125) were dropped in a secondary analysis because it is used for many indications.

The drug information is collected at each study visit by asking women to bring in their medication bottles or a pharmacy generated list of medications that they have used in the last month. Interviewers record the medications used, which are coded using the Iowa Drug Information Service system.[16] Women were not prompted specifically about sleep medications. Dosages and drug frequency were not reliably recorded and were not used for these analyses. Furthermore, over-the-counter medication use information was considered incomplete and not included in these analyses. Non-users were never users. They entered the study (index date) at visits matched in frequency distribution with the sleep medication user.

As noted, we only included new use of sleep medications. The first visit with a mention of a sleep medication

was considered the index visit. Since there are no between visit medication updates, we considered women who reported starting a sleep medication as users until their next annual SWAN visit. This design mimics an intention-to-treat analysis.

Outcomes

Three domains of sleep disturbances were self-reported at all annual SWAN visits. Women were asked to pick the answer that best describes their difficulty initiating sleep, remaining asleep and early morning awakenings during the previous 2 weeks. They used a 5-point Likert scale to report on each type of disturbance, where 1=no difficulties on any nights, 2=difficulties on less than one night per week, 3=one to two nights per week, 4=three to four nights per week and 5=five to seven nights per week.[17–19] We considered the results at 1 year to be the primary outcome and 2 years to be the secondary outcome. For the 2-year outcome, only women who had both year 1 and year 2 results were analysed.

Covariates

SWAN collects a broad range of variables at cohort entry and at each subsequent annual visit. We considered a wide range of potential covariates including demographics, comorbidities, menopausal status, body mass index (BMI), tobacco use and alcohol use. The variables unlikely to change over time (race/ethnicity and educational attainment) were collected at cohort entry and others were collected at the visit prior to the index visit. Variables were not updated after the index date. Depression was measured with the Center for Epidemiologic Studies Depression Scale,[20] anxiety with the General Anxiety Disorder-7[21] and the 36-Item Short Form Survey (SF-36) scales were used to measure pain, mental function and physical function.[21]

Statistical analyses

After assembling the analytic cohort, covariates were defined and compared across women who initiated a sleep medication and those who did not. To improve the baseline balance in characteristics, we estimated a propensity score using a logistic regression model.[22] A propensity score estimates the likelihood that women would start a sleep medication, with values ranging from 0 to 1. All covariates shown in table 1 were included in the propensity score model. We then matched women who started a medication for sleep with women who did not based on their propensity score.[23] We attempted to match two non-users for each user using a 'greedy matching' algorithm, with a maximum calliper of 0.2 of an SD of the logit of the propensity score.[24]

After matching, we examined baseline characteristics for balance using standardised mean differences (see table 1). With evidence of good balance across measured baseline characteristics, we next examined sleep disturbances at baseline and found these to be well balanced. We then examined sleep disturbance reports at 1 and 2 years, estimating means and SD, and the changes in sleep disturbance from baseline to 1 year and 1 year to 2 years. These changes were estimated and compared across medication exposure groups, using a mixed regression model. No adjustments were made, as the baseline characteristics were well balanced as noted in table 1.

Secondary analyses compared the distribution of scores on the Likert scale across medication exposures, specifically assessing for the per cent of women who reported less frequent sleep disturbance; this analysis has the benefit of not assuming a continuous or linear distribution across the five categories of the Likert scale. We also conducted a proportional odds analysis to determine if exposure to sleep medications was associated with a significant reduction in the Likert scale. Other secondary analyses used the visit before sleep medication initiation to define the baseline patient characteristics to calculate the propensity score; this analysis allows us to assess the sensitivity of the results to the timing of variable measurement. We restricted the analyses to women who reported more severe sleep disturbances at baseline, defined as a four or five on at least one sleep domain. This definition is consistent with the frequency criterion for clinically significant sleep difficulty (eg, insomnia disorder).[25 26] We compared no medication use to specific sleep medications, BZDs and selective BZD receptor agonists. Finally, we ran models adjusted for SWAN site and oestrogen replacement therapy. Such analyses retained the propensity score match.

All analyses were conducted using SAS V9.1 (Cary, North Carolina, USA). All p values were nominal and not adjusted for multiple comparisons, as these were post hoc exploratory analyses.

RESULTS

We identified 2531 potentially eligible women in SWAN who reported the severity of a sleep disturbance at some point during the 21 years of follow-up, 1995–2016 (see figure 1). We applied the exclusion criteria and found 1528 women who were analysed in the propensity score to identify potential matches. From this group, the 238 women who initiated a prescription sleep medication were significantly different than the overall group of women who did not (see online supplemental table 2). Thus, we propensity matched the 238, attempting, attempting to find two non-users for each user; we were able to match 447 women who never initiated a sleep medication during study follow-up. These 685 women were similar in characteristics to the 1846 potentially eligible women not included in the analysis (see online supplemental table 3). Hundred per cent of women included reported a sleep disturbance at some point during follow-up. At baseline, 72%–77% reported sleep disturbance.

The baseline characteristics of the women in the study cohort are shown in table 1. After propensity score matching, the women who initiated a sleep medication and those who did not were similar; all standardised mean

Table 1 Baseline demographics of women in SWAN examined in the primary cohort

	Total n=685	No sleep medication n=447	Sleep medication user n=238	
		N (%) unless noted		SMD
Age, mean (SD)	49.5 (8.5)	49.6 (8.8)	49.3 (7.7)	0.02
BMI, mean (SD)	29.1 (7.4)	29.1 (7.3)	29.2 (7.6)	0.02
Educational attainment				
High school or less	141 (20.6)	87 (19.5)	54 (22.7)	0.06
>High school	542 (79.1)	358 (80.1)	184 (77.3)	0.07
Ethnicity/Race				0.06
African-American	158 (23.1)	103 (23.0)	55 (23.1)	0.002
White	394 (57.5)	261 (58.4)	133 (55.9)	0.05
Chinese	45 (6.6)	29 (6.5)	16 (6.7)	0.009
Hispanic	25 (3.7)	15 (3.4)	10 (4.2)	0.05
Japanese	63 (9.2)	39 (8.7)	24 (10.1)	0.04
Medical insurance	660 (96.4)	430 (96.2)	230 (96.6)	0.02
Marital status				
Single	94 (13.7)	58 (13.0)	36 (15.1)	0.06
Married	451 (65.8)	305 (68.2)	146 (61.3)	0.15
Separated	19 (2.8)	9 (2.0)	10 (4.2)	0.15
Widowed	30 (4.4)	17 (3.8)	13 (5.5)	0.08
Divorced	91 (13.3)	58 (13.0)	33 (13.9)	0.03
Tobacco use				
Never	344 (50.2)	220 (49.2)	124 (52.1)	0.06
Past/Current	341 (49.8)	227 (50.8)	114 (47.9)	0.06
Alcohol use				0.05
None	294 (44.1)	193 (44.3)	101 (43.7)	0.01
<1 drink/week	167 (25.0)	117 (26.8)	50 (21.7)	0.12
1–7 drinks/week	131 (19.6)	75 (17.2)	56 (24.2)	0.17
>7 drinks/week	75 (11.2)	51 (11.7)	24 (10.4)	0.04
Depression (CES-D), mean (SD)	12.7 (10.5)	12.4 (10.3)	13.2 (10.9)	0.08
Anxiety score, mean (SD)	3.2 (2.7)	3.1 (2.8)	3.2 (2.6)	0.03
Body pain, mean (SD)	62.3 (22.5)	62.5 (22.0)	61.9 (23.3)	0.03
SF36-mental, mean (SD)	46.5 (11.3)	46.7 (11.6)	46.2 (10.8)	0.05
SF-36-physical, mean (SD)	48.1 (10.4)	48.2 (9.8)	47.9 (11.5)	0.03
Menopausal status				0.06
Unknown	85 (12.4)	52 (11.6)	33 (13.9)	0.07
Premenopausal	30 (4.6)	19 (4.3)	11 (4.6)	0.02
Early/Late peri-menopausal	246 (35.9)	162 (36.2)	84 (35.3)	0.02
Surgical menopause	30 (4.2)	20 (4.5)	10 (4.2)	0.01
Postmenopausal	294 (42.9)	194 (43.4)	100 (42.0)	0.03
Diabetes	65 (9.5)	38 (8.5)	27 (11.3)	0.10
Hypertension	316 (46.1)	201 (45.0)	115 (48.3)	0.07
Osteoarthritis	303 (44.2)	196 (43.9)	107 (45.0)	0.02
Cancer, current	21 (3.1)	16 (1.8)	5 (2.1)	0.10
Any antidepressant	22 (3.2)	6 (1.3)	16 (6.7)	0.28
Any analgesic	28 (4.1)	22 (4.9)	6 (2.5)	0.13

There are missing values for education (n=2), alcohol use (n=14) and insurance (n=25). Antidepressants include TCAs, SSRI, SNRIs and MAO inhibitors. Analgesics include opioids and non-steroidal anti-inflammatory drugs. The CES-D is a 20-item scale with a range of 0–60.[20] The anxiety score (GAD-7) is a 7-item scale with a range of 0–21.[21] The SF-36 bodily pain score includes two items with a range of 0–100; SF-36 mental component score is a 5-item scale with a range of 0–100 and SF-36 physical function is a 10-item scale with a range of 0–100.[21]

BMI, body mass index; CES-D, Centre for Epidemiologic Studies Depression Scale; GAD-7, General Anxiety Disorder-7; MAO, monoamine oxidase; SF-36, 36-Item Short Form Survey; SF-36-mental, Mental Component Score; SF-36-physical, Physical Component Score; SMD, standardised mean difference; SNRI, serotonin-norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; SWAN, Study of Women's Health Across the Nation; TCA, tricyclic antidepressant.

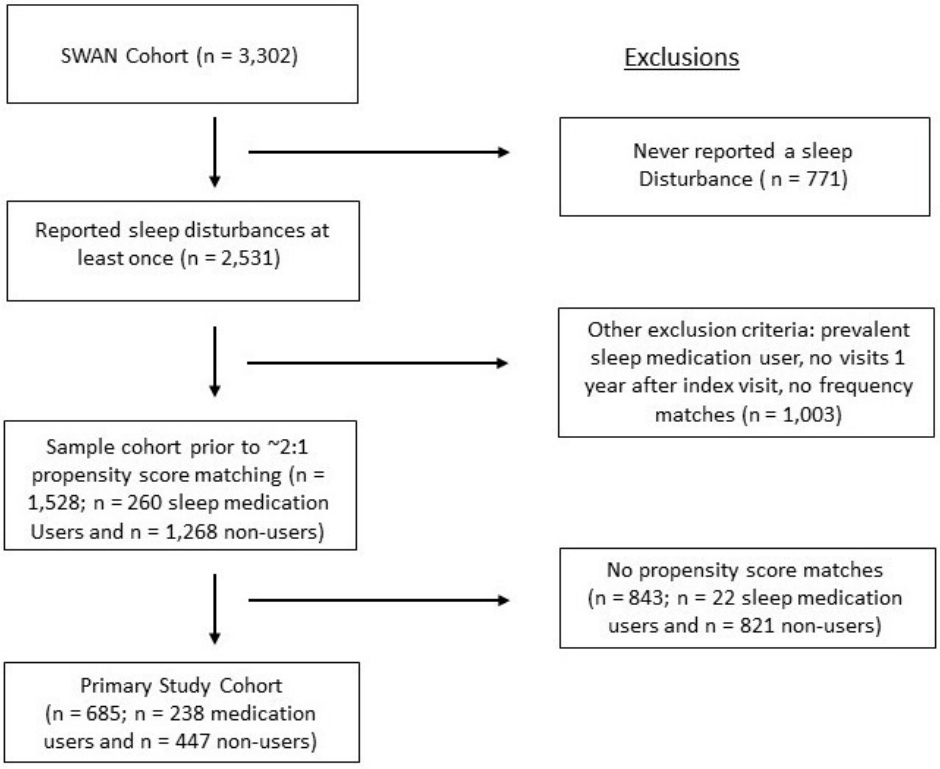

Figure 1 Assembly of the primary study cohort is demonstrated in this figure. The final study cohort was selected based on propensity score matching from the women who were potentially eligible and met selection criteria.

differences were ≤0.1, indicating successful propensity score matching. The mean age for this analytic sample was 49.5 years (SD 8.5) and their BMI was 29.1 kg/m^2 (SD 7.4). Approximately 80% had some education beyond high school. Approximately one-quarter were African-American and 57.5% were white; Hispanic, Chinese and Japanese women made up the rest of the sample. Almost all women had some medical insurance. Approximately half were current or past tobacco users and half were moderate to heavy alcohol users. Mean depression, anxiety and pain scores were similar across the groups, as were SF-36 mental and physical function scores. Menopausal status was very similar across the groups with about 36% being in the perimenopause. The range of comorbidities was typical for this population and similar across exposure groups.

At baseline, women who did and did not start a sleep medication reported very similar levels of sleep disturbance (see table 2). In both groups, women reported difficulty initiating sleep on approximately one-third of nights, waking frequently on approximately two-thirds of nights and early morning awakenings on approximately one-third of nights of the week. More than 70% of both groups reported any sleep disturbance at least 3 times weekly.

After 1 year, there were slight reductions noted in women's reports of all types of sleep disturbances, but none of the differences from baseline in either exposure

Table 2 Sleep disturbances at baseline among women in SWAN included in the primary cohort

	No sleep medication n=447	Sleep medication user n=238	SMD
Trouble initiating sleep, mean (95% CI)*	2.6 (2.5 to 2.7)	2.7 (2.5 to 2.9)	0.08
Waking frequently, mean (95% CI)*	3.7 (3.6 to 3.9)	3.8 (3.6 to 3.9)	0.03
Early morning awakening, mean (95% CI)*	2.7 (2.6 to 2.8)	2.8 (2.6 to 3.0)	0.07
Trouble initiating sleep, at least three nights per week, n (%)	137 (30.7)	82 (34.5)	0.07
Waking frequently, at least three nights per week, n (%)	291 (65.1)	158 (66.4)	0.008
Early morning awakening, at least three nights per week, n (%)	135 (30.2)	81 (34.0)	0.07
Any disturbance, at least three nights per week, n (%)	322 (72.0)	183 (76.9)	0.08

*Mean calculated based on 5-point Likert scale, where 1=no difficulties on any nights, 2=difficulties on less than one night per week, 3=one to two nights per week, 4=three to four nights per week and 5=five to seven nights per week.
SMD, standardised mean difference; SWAN, Study of Women's Health Across the Nation.

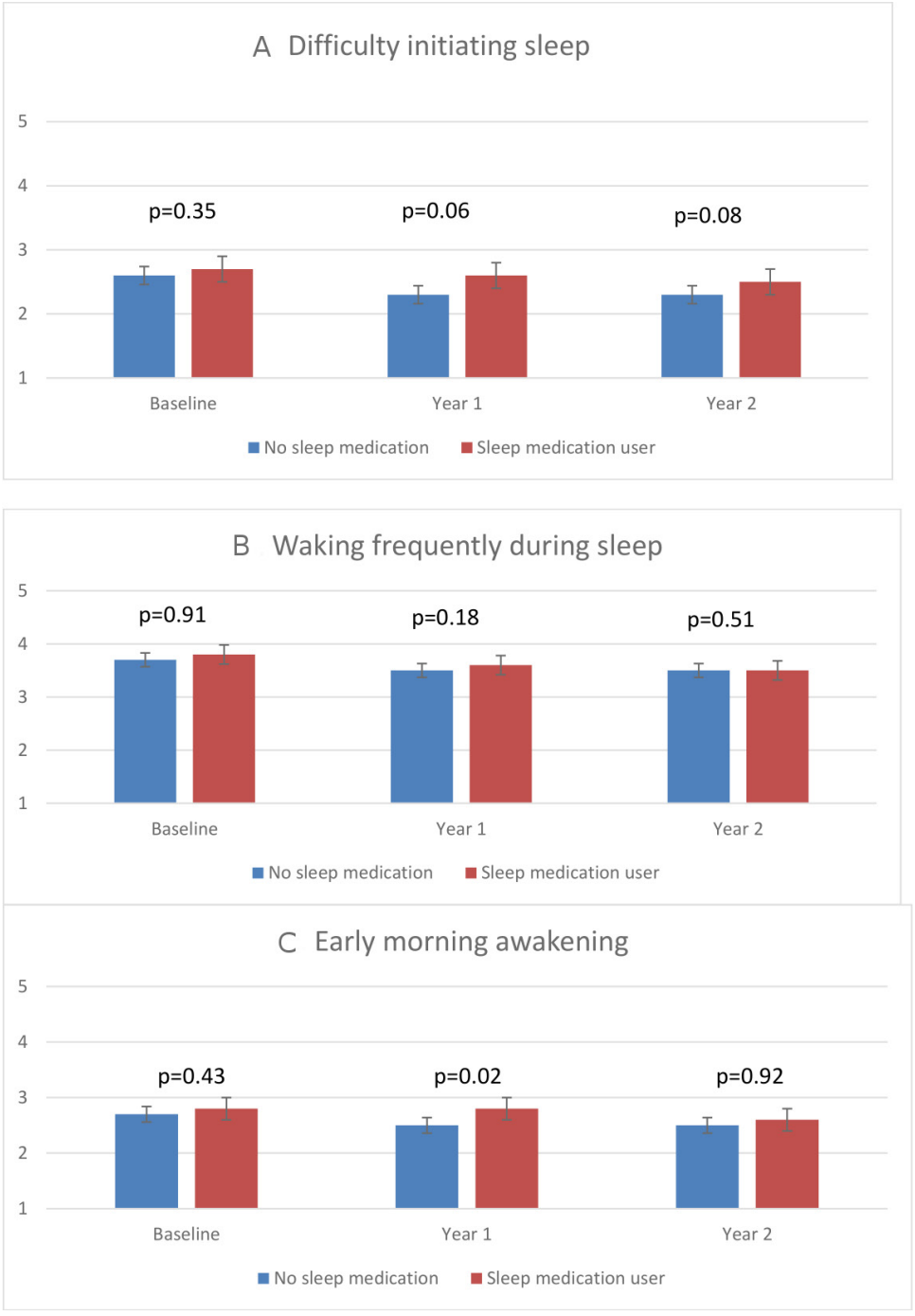

Figure 2 The three panels describe sleep disturbance ratings by medication exposure. Means were calculated based on 5-point Likert scale, where 1=no difficulties on any nights, 2=difficulties on less than one night per week, 3=one to two nights per week, 4=three to four nights per week and 5=five to seven nights per week. Error bars represent 95% CIs. P values at baseline, year 1 and year 2 comparing sleep medication users with non-users were estimated from the Wilcoxon rank-sum test. In panel A, p values for the differences between medication users and non-users for the change between baseline and 1 year=0.19; baseline and 2 years=0.55 and 1 year and 2 years=0.73. In panel B, p values for the differences between medication users and non-users for the change between baseline and 1 year=0.41; baseline and 2 years=0.98 and 1 year and 2 years=0.55. In panel C, p values for the differences between medication users and non-users for the change between baseline and 1 year=0.13; baseline and 2 years=0.46; *1 years and 2 years=0.03 (favouring non-use).

group (medication users or non-users) was statistically significant (see figure 2). One-year reports of early morning awakenings appeared to be slightly lower on the Likert scale among women not using sleep medications (mean 2.5, 95% CI 2.3 to 2.6) compared with those who did (mean 2.8, 95% CI 2.6 to 3.0; p=0.02). The secondary 2-year outcomes were similar to the 1-year results; none demonstrated statistically significant reductions in sleep disturbances among sleep medication users.

Table 3 Likert scale severity ratings of self-reported sleep disturbances from baseline to year 1 among women in SWAN who reported sleep disturbances

	Baseline visit		Visit 1 year after		Visit 2 years after	
	No sleep medication n=447	Medication users n=238	No sleep medication n=447	Medication users n=238	No sleep medication n=353	Medication users n=187
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Sleep disturbance						
Difficulty initiating sleep (per week)						
1 (no difficulty)	154 (34.5)	81 (34)	190 (42.5)	94 (39.5)	156 (44.2)	70 (37.4)
2 (≤1 night/week)	74 (16.6)	31 (13)	72 (16.1)	29 (12.2)	52 (14.7)	33 (17.6)
3 (1–2 nights/week)	81 (18.1)	44 (18.5)	82 (18.3)	37 (15.5)	70 (19.8)	32 (17.1)
4 (3–4 nights/week)	74 (16.6)	39 (16.4)	49 (11)	34 (14.3)	32 (9.1)	16 (8.6)
5 (5–7 nights/week)	63 (14.1)	43 (18.1)	54 (12.1)	44 (18.5)	43 (12.2)	36 (19.3)
Waking frequently during sleep						
1 (no difficulty)	47 (10.5)	20 (8.4)	63 (14.1)	34 (14.3)	42 (11.9)	25 (13.4)
2 (<1 night/week)	41 (9.2)	23 (9.7)	54 (12.1)	25 (10.5)	50 (14.2)	21 (11.2)
3 (1–2 nights/week)	68 (15.2)	37 (15.5)	89 (19.9)	38 (16)	78 (22.1)	36 (19.3)
4 (3–4 nights/week)	118 (26.4)	69 (29)	93 (20.8)	47 (19.7)	70 (19.8)	40 (21.4)
5 (5–7 nights/week)	173 (38.7)	89 (37.4)	148 (33.1)	94 (39.5)	113 (32)	65 (34.8)
Early morning awakening						
1 (no difficulty)	127 (28.4)	69 (29)	171 (38.3)	72 (30.3)	122 (34.6)	70 (37.4)
2 (<1 night/week)	83 (18.6)	37 (15.5)	82 (18.3)	49 (20.6)	72 (20.4)	30 (16)
3 (1–2 nights/week)	102 (22.8)	51 (21.4)	67 (15)	35 (14.7)	67 (19)	34 (18.2)
4 (3–4 nights/week)	76 (17)	39 (16.4)	66 (14.8)	30 (12.6)	41 (11.6)	20 (10.7)
5 (5–7 nights/week)	59 (13.2)	42 (17.6)	61 (13.6)	52 (21.8)	51 (14.4)	33 (17.6)
Any complaint of 3 or more times per week*						
Yes	322 (72)	183 (76.9)	273 (61.1)	159 (66.8)	203 (57.5)	122 (65.2)

Means calculated based on 5-point Likert scale, where 1=no difficulties on any nights, 2=difficulties on less than one night per week, 3=one to two nights per week, 4=three to four nights per week and 5=five to seven nights per week.
SWAN, Study of Women's Health Across the Nation.

Several secondary analyses were pursued. First, we examined the distribution of Likert scores at baseline and 1 year of follow-up in the two groups (see table 3). The distributions among medication users and non-users were similar at baseline and follow-up (all p values >0.10). We also examined whether the results differed by type of sleep medication, BZD versus selective BZD receptor agonists and other hypnotics (see table 4); no differences were observed in the change from baseline to 1 year for either sleep medication group compared with medication non-users. The BZD group was further examined after removing lorazepam, and we found similar results for all types of sleep disturbances. We also re-ran the analyses with the baseline characteristics defined at the visit prior to the start of medications to assess how sensitive the results were to possible imprecision in the timing of variable measurement. The results showed small improvements in early morning awakenings among the sleep medication group (see online supplemental table 4). Additional sensitivity analyses retained the five-level categorical Likert scale as the primary outcome and proportional odds analyses gave similar negative results (see table 3 and online supplemental table 5); all proportional odds assumptions were met. In analyses that only included the women reporting clinically significant weekly frequency of sleep disturbances at baseline (4 or 5 on the Likert scale), no differences were found between sleep medication users and non-users (see online supplemental figure 2). Finally, analyses that also included site and oestrogen use gave similar results (see online supplemental table 6).

DISCUSSION

Sleep difficulties are common.[1 27] Not surprisingly, the use of sleep medications has also grown over the last two decades.[2] These agents have a range of safety concerns[5] and recent reports describe substantial driving impairments.[28] Most data regarding their efficacy derive from short-term studies (ie, 2–12 weeks), but these agents appear to be used over the long-term by many patients. In this analysis of the long-term impact of sleep medications in a large longitudinal cohort of well-characterised middle-aged community-dwelling women with sleep

Table 4 Change in severity of self-reported sleep disturbances from baseline to year 1 among women in SWAN who reported sleep disturbances, by medication type

	Baseline visit		Visit 1 year after		
	No sleep medications n=447	BZD users n=87	No sleep medications n=447	BZD users n=87	
	Mean (95% CI)				P value*
Difficulty initiating sleep	2.6 (2.5 to 2.7)	2.2 (2.5 to 3.2)	2.3 (2.2 to 2.5)	2.6 (2.3 to 2.9)	0.71
Waking frequently during sleep	3.7 (3.6 to 3.8)	3.8 (3.5 to 4.1)	3.5 (3.3 to 3.6)	3.3 (3.0 to 3.6)	0.24
Early morning awakening	2.7 (2.6 to 2.8)	2.6 (2.3 to 2.9)	2.5 (2.3 to 2.6)	2.6 (2.3 to 2.9)	0.17
	No sleep medications n=447	Z-drugs+other hypnotics n=151	No sleep medications n=447	Z-drugs+other hypnotics n=151	
Difficulty initiating sleep	2.6 (2.5 to 2.7)	2.7 (2.4 to 3.0)	2.3 (2.2 to 2.4)	2.6 (2.3 to 2.8)	0.12
Waking frequently during sleep	3.7 (3.6 to 3.8)	3.8 (3.6 to 4.0)	3.5 (3.3 to 3.6)	3.8 (3.6 to 4.0)	0.05
Early morning awakening	2.7 (2.5 to 2.8)	2.9 (2.7 to 3.1)	2.5 (2.3 to 2.6)	2.8 (2.6 to 3.0)	0.28

Z-drugs (selective BZD receptor agonists) include zolpidem, zaleplon and eszopiclone. Means calculated based on 5-point Likert scale, where 1=no difficulties on any nights, 2=difficulties on less than one night per week, 3=one to two nights per week, 4=three to four nights per week and 5=five to seven nights per week.
*P values reflect the differences between the sleep medication users and non-users in the change in severity of disturbances between baseline and year 1.
BZD, benzodiazepine; SWAN, Study of Women's Health Across the Nation.

disturbances, sleep medication use was not associated with reduced sleep disturbances.

When physicians or other clinicians prescribe these medicines, they often begin with short-term prescriptions, but many patients receiving these prescriptions become long-term users.[9] In the SWAN cohort, 37% of women starting a medication for sleep report using a sleep medication 1 year later. While there are good data from RCTs that these medications improve sleep disturbances in the short term,[8] the results we present here represent some of the only data on these medications' long-term impact on sleep. The lack of benefit observed in the current study suggests that when physicians begin prescribing these medicines they should discuss with patients that many patients continue them long-term, and that there is scant evidence demonstrating benefit to using these medicines beyond several months.[6 7] In the study cohort, approximately half of the women were current or past tobacco users and 20% were moderate to heavy alcohol users. This was higher than expected and may reflect the demographic of women who endorse having a sleep disturbance.

A broader issue raised by this example is how clinicians should consider prescribing medications when their expected use differs substantially from the RCT evidence. Without evidence from RCTs demonstrating the benefit of a given type of drug in a given patient population using the drug for a similar duration, clinicians lack the necessary information to prescribe appropriately. Real-world data, or data from observational cohorts such as what we present here, provide important opportunities for looking at the way drugs may actually be used in typical practice. There has been an increasing appreciation for the use of observational data analysed appropriately to complement RCTs.[10] The Food and Drug Administration has published a framework for generating evidence from real-world observational data sets,[29] with the hope that such analyses will allow clinicians to better understand the benefits and risks of drugs in typical practice.

We used rigorous epidemiological methods and analysed a well-characterised cohort of women, but as with all observational studies there are limitations to recognise. The use of sleep medications was not randomised. Thus, even though the propensity score matched cohorts were very similar, there may be unmeasured confounding not accounted for in the analyses. These analyses were not predefined prior to establishing the SWAN cohort and should be considered post hoc and exploratory. Medication use was collected only at annual or biennial study visits, and there may have been intermittent use or non-adherence between visits. This is a limitation of many retrospective cohort medication analyses and limits the inferences that can be drawn. In the primary 1-year analysis, women were required to report use of a sleep medication at the subsequent annual visit in the new initiator group and to not report a sleep medication in the non-user group. In the secondary 2-year analysis, women who remained on drug accrued no benefit compared with women who never used a sleep medication. We did not update covariates in the 2-year analysis.

Sleep disturbances were self-reported, without any objective measures of sleep. This may have introduced misclassification, however the outcomes were self-reported among both groups of women, limiting any potential bias. The outcome measure we used for sleep disturbances has been validated in prior studies,[17 18] but never in SWAN participants. The five-level categorical Likert scale was primarily analysed as a continuous variable in the mixed

regression models, however analyses that retained the five categories gave similar negative results (see table 3 and online supplemental table 4). We do not have measures of daytime consequences in this dataset. It is also possible that sleep medications may have helped in the short-term, that is, at 8 or 12 weeks. Women only reported medication use and sleep disturbances at annual visits and thus interim outcomes (ie, at 6-month intervals) and intermittent medication use are not available for analysis. We did not include over-the-counter medication use and thus some non-users may actually have been using an over-the-counter hypnotic. We know that 11% of the women in this study reported use of an over-the-counter hypnotic at the baseline visit; slightly more women in the user group reported such use compared with the non-user group. Finally, some prescription sleep medications can be used for multiple indications, regardless of the prescriber's knowledge.

In addition to these limitations, several strengths of this study should be described. We examined a well-characterised cohort of women during a high-risk period for sleep disturbance. It is known that women going through the midlife often note sleep disturbances.[30] As well, we studied women of several races and ethnicities, enhancing the generalisability of the results. The study design also allowed us to examine a well-balanced cohort with very similar identical baseline features after propensity score matching. However, unmeasured or residual confounding cannot be ruled out.

In conclusion, sleep disturbances are common and increasing in prevalence. The use of sleep medications has grown, and they are often used over a long period, despite the relative lack of evidence from RCTs. The current observational study does not support use of sleep medications over the long term, as there were no self-reported differences at 1 or 2 years of follow-up comparing sleep medication users with non-users. While we used rigorous epidemiological methods, the findings reported herein are based on a non-randomised observational dataset and must be seen in that light. It is also important to note that neither group reported more severe sleep disturbances over the study follow-up. Most patients, if not all, should have received cognitive behavioural therapy.[31] While some small percentage of patients with sleep disturbances may receive benefit from using these medications over several years, the lack of benefit associated with use of sleep medications in the population studied after 1 and 2 years should help inform clinicians and patients considering initiating pharmacological treatment for midlife women who have sleep complaints.

Author affiliations
[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA
[2]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[3]Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[4]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA
[5]Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
[6]Psychiatry, Harvard Univerisity, Boston, Massachusetts, USA
[7]Department of Psychiatry, Rush Medical College of Rush University, Chicago, Illinois, USA

Twitter Daniel H Solomon @DanielHSolomon

Contributors DHS: design, analysis, drafting and revising manuscript. KR: analysis and revising manuscript. LH: design, analysis and revising manuscript. JF: data collection, design and revising manuscript. PL: analysis and revising manuscript. HJ: design and revising manuscript. HMK: data collection, design and revising manuscript.

Funding The Study of Women's Health Across the Nation (SWAN) has grant support from the National Institutes of Health (NIH), DHHS, through the National Institute on Ageing (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women's Health (ORWH) (Grants U01NR004061; U01AG012505, U01AG012535, U01AG012531, U01AG012539, U01AG012546, U01AG012553, U01AG012554, U01AG012495). Clinical Centres: University of Michigan, Ann Arbour—Siobán Harlow, PI 2011–present, MaryFran Sowers, PI 1994–2011; Massachusetts General Hospital, Boston, Massachusetts—Joel Finkelstein, PI 1999–present; Robert Neer, PI 1994–1999; Rush University, Rush University Medical Centre, Chicago, Illinois—Howard Kravitz, PI 2009–present; Lynda Powell, PI 1994–2009; University of California, Davis/Kaiser—Ellen Gold, PI; University of California, Los Angeles—Gail Greendale, PI; Albert Einstein College of Medicine, Bronx, New York—Carol Derby, PI 2011–present, Rachel Wildman, PI 2010–2011; Nanette Santoro, PI 2004–2010; University of Medicine and Dentistry—New Jersey Medical School, Newark—Gerson Weiss, PI 1994–2004 and the University of Pittsburgh, Pittsburgh, Pennsylvania—Karen Matthews, PI. NIH Programme Office: National Institute on Ageing, Bethesda, Maryland—Chhanda Dutta 2016–present; Winifred Rossi 2012–2016; Sherry Sherman 1994–2012; Marcia Ory 1994–2001; National Institute of Nursing Research, Bethesda, Maryland—Programme Officers. Central Laboratory: University of Michigan, Ann Arbour—Daniel McConnell (Central Ligand Assay Satellite Services). Coordinating Centre: University of Pittsburgh, Pittsburgh, Pennsylvania—Maria Mori Brooks, PI 2012–present; Kim Sutton-Tyrrell, PI 2001–2012; New England Research Institutes, Watertown, Massachusetts—Sonja McKinlay, PI 1995–2001. Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair.

Disclaimer The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIA, NINR, ORWH or the NIH.

Competing interests DHS also receives support from NIH-P30-AR072577. He has received salary support from research grants to Brigham and Women's Hospital for unrelated work from AbbVie, Amgen, Corrona, Genentech and Pfizer.

Patient consent for publication Not required.

Ethics approval This protocol was reviewed and approved at each participating SWAN site: University of Pittsburgh—REN15070236/IRB0709006; Massachusetts General Hospital—1999P006353; University of Michigan—00000245; Albert Einstein College of Medicine—2005-012; Rush University Medical Centre—13021201-IRB01-AM04; University of California, Davis—260339-17; UCLA—11-002274-AM-00009.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. This is an NIH-funded study and data are accessible through appropriate channels.

ORCID iD

Daniel H Solomon http://orcid.org/0000-0001-8202-5428

REFERENCES

1 Bertisch SM, Herzig SJ, Winkelman JW, *et al*. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep* 2014;37:343–9.
2 Kaufmann CN, Spira AP, Alexander GC, *et al*. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. *Pharmacoepidemiol Drug Saf* 2016;25:637–45.
3 Ford ES, Wheaton AG, Cunningham TJ, *et al*. Trends in outpatient visits for insomnia, sleep apnea, and prescriptions for sleep medications among US adults: findings from the National ambulatory medical care survey 1999-2010. *Sleep* 2014;37:1283–93.
4 Bhaskar S, Hemavathy D, Prasad S. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. *J Family Med Prim Care* 2016;5:780–4.
5 Schroeck JL, Ford J, Conway EL, *et al*. Review of safety and efficacy of sleep medicines in older adults. *Clin Ther* 2016;38:2340–72.
6 Krystal AD, Walsh JK, Laska E, *et al*. Sustained efficacy of eszopiclone over 6 months of nightly treatment: results of a randomized, double-blind, placebo-controlled study in adults with chronic insomnia. *Sleep* 2003;26:793–9.
7 Randall S, Roehrs TA, Roth T. Efficacy of eight months of nightly zolpidem: a prospective placebo-controlled study. *Sleep* 2012;35:1551–7.
8 Sateia MJ, Buysse DJ, Krystal AD, *et al*. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of sleep medicine clinical practice guideline. *J Clin Sleep Med* 2017;13:307–49.
9 Jaussent I, Ancelin M-L, Berr C, *et al*. Hypnotics and mortality in an elderly general population: a 12-year prospective study. *BMC Med* 2013;11:212.
10 Franklin JM, Glynn RJ, Martin D, *et al*. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther* 2019;105:867–77.
11 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
12 Schneeweiss S, Rassen JA, Brown JS, *et al*. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med* 2019;170:398–406.
13 Matthews KA, Crawford SL, Chae CU, *et al*. Are changes in cardiovascular disease risk factors in midlife women due to chronological aging or to the menopausal transition? *J Am Coll Cardiol* 2009;54:2366–73.
14 Karlamangla AS, Singer BH, Williams DR, *et al*. Impact of socioeconomic status on longitudinal accumulation of cardiovascular risk in young adults: the cardia study (USA). *Soc Sci Med* 2005;60:999–1015.
15 Sowers M, Crawford S, Sternfeld B. *Design, survey, sampling and recruitment methods of Swan: a multi-center, multi-ethnic, community based cohort study of women and the menopausal transition*. San Diego: Academic Press, 2000.
16 Milne A. A pilot study to examine the comparative usefulness of the Iowa drug information service (IDIS) and Medline in meeting the demands of a pharmacy-based drug-information service. *J Clin Pharm Ther* 1977;2:227–37.
17 Jenkins CD, Stanton BA, Niemcryk SJ, *et al*. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol* 1988;41:313–21.
18 Levine DW, Dailey ME, Rockhill B, *et al*. Validation of the women's health Initiative insomnia rating scale in a multicenter controlled clinical trial. *Psychosom Med* 2005;67:98–104.
19 Levine DW, Kripke DF, Kaplan RM, *et al*. Reliability and validity of the women's health Initiative insomnia rating scale. *Psychol Assess* 2003;15:137–48.
20 Carleton RN, Thibodeau MA, Teale MJN, *et al*. The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. *PLoS One* 2013;8:e58067.
21 Ware JE. Sf-36 health survey update. *Spine* 2000;25:3130–9.
22 Brookhart MA, Schneeweiss S, Rothman KJ, *et al*. Variable selection for propensity score models. *Am J Epidemiol* 2006;163:1149–56.
23 Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biom J* 2009;51:171–84.
24 Parsons LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques, 2014. Available: www2.sas.com/proceedings/sugi26/p214-26.pdf
25 Association AP. *Diagnostic and statistical manual of mental disorders*. 5 edn. Arlington, VA, 2013.
26 Medicine AAoS. American Academy of Sleep Medicine. In: *International classification of sleep disorders*. 3 edn. Darien, IL: American Academy of Sleep Medicine, 2014.
27 Kaufman DW, Kelly JP, Rosenberg L, *et al*. Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey. *JAMA* 2002;287:337–44.
28 Verster JC, Mooren L, Bervoets AC, *et al*. Highway driving safety the day after using sleep medication: the direction of lapses and excursions out-of-lane in drowsy drivers. *J Sleep Res* 2018;27:e12622.
29 Administration FaD. *Framework for FDA's real-world evidence program*, 2018.
30 Kravitz HM, Janssen I, Bromberger JT, *et al*. Sleep trajectories before and after the final menstrual period in the study of women's health across the nation (Swan). *Curr Sleep Med Rep* 2017;3:235–50.
31 Medalie L, Cifu AS. Management of chronic insomnia disorder in adults. *JAMA* 2017;317:762–3.