BMJ Open

The Real-World Effectiveness of Prescription Medications Among Midlife Women with Sleep Disturbances During Two Years of Follow-up: A SWAN Pharmacoepidemiology Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045074
Article Type:	Original research
Date Submitted by the Author:	22-Sep-2020
Complete List of Authors:	Solomon, Daniel; Brigham and Women's Hospital, Division of Rheumatology, Immunology and Allergy; Brigham and Women's Hospital, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine Ruppert, Kristine; University of Pittsburgh Habel, Laurel; Kaiser Permanente Northern California, Division of Research Finkelstein, Joel; Massachusetts General Hospital Lian, Pam; University of Pittsburgh Joffe, Hadine; Harvard Univerisity, Psychiatry Kravitz, Howard M.; Rush Medical College of Rush University Department of Anesthesiology
Keywords:	EPIDEMIOLOGY, SLEEP MEDICINE, THERAPEUTICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

The Real-World Effectiveness of Prescription Medications Among Midlife Women with Sleep Disturbances During Two Years of Follow-up: A SWAN Pharmacoepidemiology Study

Daniel H. Solomon (1), Kristine Ruppert (2), Laurel Habel (4), Joel Finkelstein (5), Pam Lian (3), Hadine Joffe (2), Howard M. Kravitz (6)

From: (1) Division of Rheumatology and Pharmacoepidemiology, Department of Medicine, Brigham and Women's Hospital, Boston MA; (2) Department of Epidemiology, University of Pittsburgh, Pittsburgh PA; (3) Division of Research, Kaiser Permanente Northern California, Oakland CA; (4) Endocrine Unit, Massachusetts General Hospital, Boston MA; (5) Department of Psychiatry, and Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital, Boston MA; (6) Department of Psychiatry and Department of Preventive Medicine, Rush University Medical Center, Chicago, IL

Support: 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 Aging (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). 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. Clinical Centers: University of Michigan, Ann Arbor – Siobán Harlow, PI 2011 – present, MaryFran Sowers, PI 1994-2011; Massachusetts General Hospital, Boston, MA – Joel Finkelstein, PI 1999 – present; Robert Neer, PI 1994 – 1999; Rush University, Rush University Medical Center, Chicago, IL – 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, NY – 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, PA – Karen Matthews, PI. NIH Program Office: National Institute on Aging, Bethesda, MD – Chhanda Dutta 2016- present; Winifred Rossi 2012–2016; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD – Program Officers. Central Laboratory: University of Michigan, Ann Arbor – Daniel McConnell (Central Ligand Assay Satellite Services). Coordinating Center: University of Pittsburgh, Pittsburgh, PA – Maria Mori Brooks, PI 2012 - present; Kim Sutton-Tyrrell, PI 2001 – 2012; New England Research Institutes, Watertown, MA - Sonja McKinlay, PI 1995 – 2001. Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair

Potential Conflicts: Dr. Solomon 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.

Correspondence: Daniel H. Solomon, Division of Rheumatology, Brigham and Women's Hospital, 60 Fenwood Road, Boston MA, 02115. T: 617-732-5356. F: 617-732-5505. E: dsolomon@bwh.harvard.edu

Words: 3389; Tables: 4, Figures: 2; Citations: 31

Supplements: Yes

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 US.

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 two weeks – difficulty initiating asleep, 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 one-year (primary outcome) and two-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 asleep of 2.7 (SD 1.5), waking frequently 3.8 (SD 1.3), and early morning awakening 2.8 (SD 1.5); non-users ratings were 2.6 (SD 1.5), 3.7 (SD 1.3), and 2.7 (SD 1.4), respectively. After one year, ratings for medication users were 2.6 (SD 1.6) for initiating asleep, 3.6 (SD 1.5) for waking frequently, and 2.8 (SD 1.5) for early morning awakening; for non-users, the mean ratings were 2.3 (SD 1.4), 3.5 (SD 1.4), and 2.5 (SD 1.5), respectively. None of the one-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 one and two years. The potential benefits of long-term use of sleep medication should be re-examined.

Article summary: Strengths and limitations of this study:

- Little is known about the long-term effectiveness of medications used for sleep.
- 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.
- No clear differences were observed in sleep difficulties at one and two years of follow-up between women who did and did not start 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 United States 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 pharmacologic treatment options have been developed for sleep.

Prescription sleep medications consist of benzodiazepines (BZDs) and non-benzodiazepine hypnotics (non-BZDs). The non-BZDs include zolpidem, zaleplon, eszopiclone, and other agents mostly used off-label to promote sleep through a variety of other mechanisms. Randomized controlled trials 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, effectiveness data would be useful for patients and clinicians if it included sleep medications used over several months in populations of typical patients with sleep disturbances; we found no such studies in the literature.

There has been increased interest in using non-randomized designs to test the real-world effectiveness of drugs.[10] We assessed the effectiveness of sleep medications among a large and diverse cohort of mid-life 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 Hernan and Robins.[11] In this study paradigm, a target randomized controlled trial 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 **Supplementary 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. Further, we described the study design using standardized illustrations as suggested by Schneeweiss and colleagues (see **Supplementary 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 multicenter, multi-ethnic/multi-racial 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 to 52 years; intact uterus and at least one ovary, not using sex steroid hormones or pregnant, breastfeeding or lactating at enrollment or within the previous three 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 3 nights per week during a two-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 two groups of medications: BZDs and non-BZDs. The full list of medications considered included the following BZDs: estazolam, flurazepam, lorazepam, temazepam, and triazolam; the following non-BZDs: 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. Further, over-the-counter medication use information was considered incomplete and not included in these analyses. Non-users were not included if they became users at a later visit.

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 two weeks. They used a five-point Likert scale to report on each type of disturbance, where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.[17-19] We considered the results at one-year to be the primary outcome and two-years to be the secondary outcome. For the two-year outcome, only women who had both year one and year two results were analyzed.

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 CES-D,[20] and the 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 zero to one. 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 2 non-users for each user using a "greedy matching" algorithm, with a maximum caliper of 0.2.[24]

After matching, we examined baseline characteristics for balance using standardized 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 one- and two-years, estimating means and standard deviations, and the changes in sleep disturbance from baseline to one year and one year to two years. These changes were estimated and compared across medication exposure groups, using a mixed model regression. 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 percent 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. As well, we 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. As well, we restricted the analyses to women who reported more severe sleep disturbances at baseline, defined as a 4 or 5 on at least one sleep domain. This definition is consistent with the frequency criterion for clinically significant sleep difficulty (e.g., insomnia disorder).[25,26] Such analyses retained the propensity score match.

All analyses were conducted using SAS (Cary NC). All p-values were nominal and not adjusted for multiple comparisons, as these were post-hoc exploratory analyses.

RESULTS

We identified 2,531 potentially eligible women in SWAN who reported the severity of a sleep disturbance at some point during the twenty-one years of follow-up, 1995-2016 (see **Figure 1**). We applied the exclusion criteria and found 1,528 women who were analyzed in the propensity score to identify potential matches. From this group, 238 women who initiated a prescription sleep medication could be matched with 447 women who never initiated a sleep medication during study follow-up. These 685 women were similar in characteristics to the 1,846 potentially eligible women not included in the analysis (see **Supplementary Table 2**).

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 standardized mean 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/m2 (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 peri-menopause. 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 one 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 group (medication users or non-users) were 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, SD 1.5) compared to those who did (mean 2.8, SD 1.5; p = 0.02). The secondary two-year outcomes were similar to the one-year results; none demonstrated statistically significant reductions in sleep disturbances among sleep medication users.

Several secondary analyses were pursued. First, we examined the distribution of Likert scores at baseline and one 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 non-BZD (see **Table 4**); no differences were observed in the change from baseline to one 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 **Supplementary Table 3**). Additional sensitivity analyses retained the five-level categorical Likert scale as the primary outcome and analyses gave similar negative results (see **Table 3 and Supplementary Table 4**). Finally, 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 **Supplementary Figure 2**).

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 (i.e., 2-12 weeks), but these agents appear to be used over the long-term by many patients. In this analysis of the long-term effectiveness in a large "real-world" longitudinal cohort of well-characterized middle-aged community-dwelling women who

self-reported their sleep disturbances and sleep medication use, sleep disturbances did not improve over one or two years among those who started sleep medications compared with women who did not.

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 one year later. While there are good data from randomized controlled trials 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 effectiveness. 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 twenty percent 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 randomized controlled trial (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 analyzed appropriately to complement randomized trials.[10] The FDA 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 epidemiologic methods and analyzed a well characterized cohort of women, but as with all observational studies there are limitations to recognize. The use of sleep medications was not randomized. 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 pre-defined 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 one-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 two-year analysis, women who remained on drug accrued no benefit compared with women who never used a sleep medication.

Sleep disturbances were self-reported, without any objective measures of sleep. This may have introduced mis-classification, 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 analyzed as a continuous variable in the mixed regression models, however analyses that retained the five categories gave similar negative results (see Table 3 and Supplementary 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, i.e., at 8 or 12 weeks. Women only reported medication use and sleep disturbances at annual visits and thus interim outcomes (i.e., at six 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 characterized 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 generalizability 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 randomized controlled clinical trials. The current observational study does not support use of sleep medications over the long-term, as there were no self-reported differences at one- or two-years of follow-up comparing sleep medication users to non-users. While we used rigorous epidemiologic methods, the findings reported herein are based on a non-randomized 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 behavioral therapy.[31] While some small percentage of patients with sleep disturbances may receive benefit from long-term use of medications, the lack of benefit associated with use of sleep medications in the population studied after one- and two-years should help inform clinicians and patients considering initiating pharmacologic treatment for midlife women who have sleep complaints.

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	SMD
		N (%) unless noted		
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
SF36-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
Pre-menopausal	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
Post-menopausal	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

Abbreviations: SMD, standardized mean difference; CES-D, Center for Epidemiologic Studies Depression Scale; BMI, Body Mass Index; SF36 Mental, Mental Component Score; and SF36 Physical, Physical Component Score. 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 nonsteroidal anti-inflammatory drugs.

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 (SD)*	2.6 (1.5)	2.7 (1.5)	0.08
Waking frequently, mean (SD)*	3.7 (1.3)	3.8 (1.3)	0.03
Early morning awakening, mean (SD)*	2.7 (1.4)	2.8 (1.5)	0.07
Trouble initiating sleep, at least 3 nights per week, n (%)	137 (30.7)	82 (34.5)	0.07
Waking frequently, at least 3 nights per week, n (%)	291 (65.1)	158 (66.4)	0.008
Early morning awakening, at least 3 nights per week, n (%)	135 (30.2)	81 (34.0)	0.07
Any disturbance, at least 3 nights per week, n (%)	322 (72.0)	183 (76.9)	0.08

Abbreviations: SD, standard deviation; SMD, standardized mean difference. *Mean calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

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	
Sleep Disturbance	N	%	n	%	n	%	n	%	n	%	n	%
Difficulty initiating sleep (per week)												
1 (no difficulty)	154	34.5%	81	34.0%	190	42.5%	94	39.5%	156	44.2%	70	37.4%
2 (= <1 night/week)	74	16.6%	31	13.0%	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 (3-4nights/week)	74	16.6%	39	16.4%	49	11.0%	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	16.0%	78	22.1%	36	19.3%
4 (3-4 nights/week)	118	26.4%	69	29.0%	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	32.0%	65	34.8%
Early morning awakening												
1 (no difficulty)	127	28.4%	69	29.0%	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	16.0%
3 (1-2 nights/week)	102	22.8%	51	21.4%	67	15.0%	35	14.7%	67	19.0%	34	18.2%
4 (3-4 nights/week)	76	17.0%	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	72.0%	183	76.9%	273	61.1%	159	66.8%	203	57.5%	122	65.2%

Abbreviations: SD, standard deviation. Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

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	P-value*
Difficulty initiating sleep, mean (SD)	2.6 (1.5)	2.2 (1.6)	2.3 (1.4)	2.6 (1.6)	0.71
Waking frequently during sleep, mean (SD)	3.7 (1.3)	3.8 (1.3)	3.5 (1.4)	3.3 (1.4)	0.24
Early morning awakening, mean (SD)	2.7 (1.4)	2.6 (1.5)	2.5 (1.5)	2.6 (1.6)	0.17
	No Sleep Medications n=447	**Non-BZD** n=151	**No Sleep Medications** n=447	**Non-BZD** n=151	
Difficulty initiating sleep, mean (SD)	2.6 (1.5)	2.7 (1.5)	2.3 (1.4)	2.6 (1.6)	0.12
Waking frequently during sleep, mean (SD)	3.7 (1.3)	3.8 (1.2)	3.5 (1.4)	3.8 (1.4)	0.05
Early morning awakening, mean (SD)	2.7 (1.4)	2.9 (1.5)	2.5 (1.5)	2.8 (1.5)	0.28

Abbreviations: SD, standard deviation; BZD, benzodiazepine; non-BZD, non-benzodiazepine. Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 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 one.

Contributorship statement:

Daniel H. Solomon: Design, analysis, drafting and revising manuscript.
Kristine Ruppert: Analysis and revising manuscript.
Laurel Habel: Design, analysis and revising manuscript.
Joel Finkelstein: Data collection, design, and revising manuscript.
Pam Lian: Analysis and revising manuscript.
Hadine Joffe: Design and revising manuscript.
Howard M. Kravitz: Data collection, design and revising manuscript.

Competing interests: Dr. Solomon 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.

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 Aging (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). 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.

Data sharing statement: Data from SWAN are available for qualified researchers.

REFERENCES

1. Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep.* 2014;37(2):343-349.
2. Kaufmann CN, Spira AP, Alexander GC, Rutkow L, Mojtabai R. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. *Pharmacoepidemiology and drug safety.* 2016;25(6):637-645.
3. Ford ES, Wheaton AG, Cunningham TJ, Giles WH, Chapman DP, Croft JB. 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(8):1283-1293.
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(4):780-784.
5. Schroeck JL, Ford J, Conway EL, et al. Review of Safety and Efficacy of Sleep Medicines in Older Adults. *Clin Ther.* 2016;38(11):2340-2372.
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(7):793-799.
7. Randall S, Roehrs TA, Roth T. Efficacy of eight months of nightly zolpidem: a prospective placebo-controlled study. *Sleep.* 2012;35(11):1551-1557.
8. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. 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(2):307-349.
9. Jaussent I, Ancelin ML, 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, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clinical pharmacology and therapeutics.* 2019;105(4):867-877.
11. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764.
12. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. *Ann Intern Med.* 2019;170(6):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(25):2366-2373.
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(5):999-1015.
15. Sowers M, Crawford S, Sternfeld B, et al. *Design, survey, sampling and recruitment methods of SWAN: a multi-center, multi-ethnic, community based cohort study of women and the menopausal transitioin. 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. *Journal of Clinical Pharmacy and Therapeutics.* 2008;2:227-237.
17. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol.* 1988;41(4):313-321.

18. Levine DW, Dailey ME, Rockhill B, Tipping D, Naughton MJ, Shumaker SA. Validation of the Women's Health Initiative Insomnia Rating Scale in a multicenter controlled clinical trial. *Psychosom Med.* 2005;67(1):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(2):137-148.
20. Carleton RN, Thibodeau MA, Teale MJ, 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(3):e58067.
21. Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976).* 2000;25(24):3130-3139.
22. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *American Journal of Epidemiology.* 2006;163(12):1149-1156.
23. Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biometrical journal.* 2009;51(1):171-184.
24. Parsons LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques. SAS. www2.sas.com/proceedings/sugi26/p214-26.pdf. Accessed 2014.
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. 5th ed (DSM-5).* Arlington, VA2013.
26. American Academy of Sleep Medicine. *International Classification of Sleep Disorders. 3rd ed.* Darien, IL: American Academy of Sleep Medicine; 2014.
27. Kaufman DW, Kelly JP, Rosenberg L, Anderson TE, Mitchell AA. Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey.[see comment]. *JAMA.* 2002;287(3):337-344.
28. Verster JC, Mooren L, Bervoets AC, Roth T. 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(3):e12622.
29. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. In: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(3):235-250.
31. Medalie L, Cifu AS. Management of Chronic Insomnia Disorder in Adults. *JAMA.* 2017;317(7):762-763.

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.

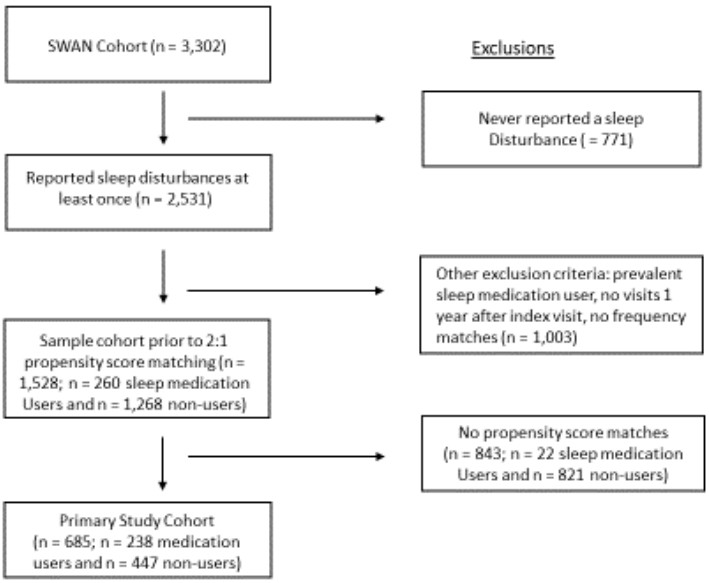
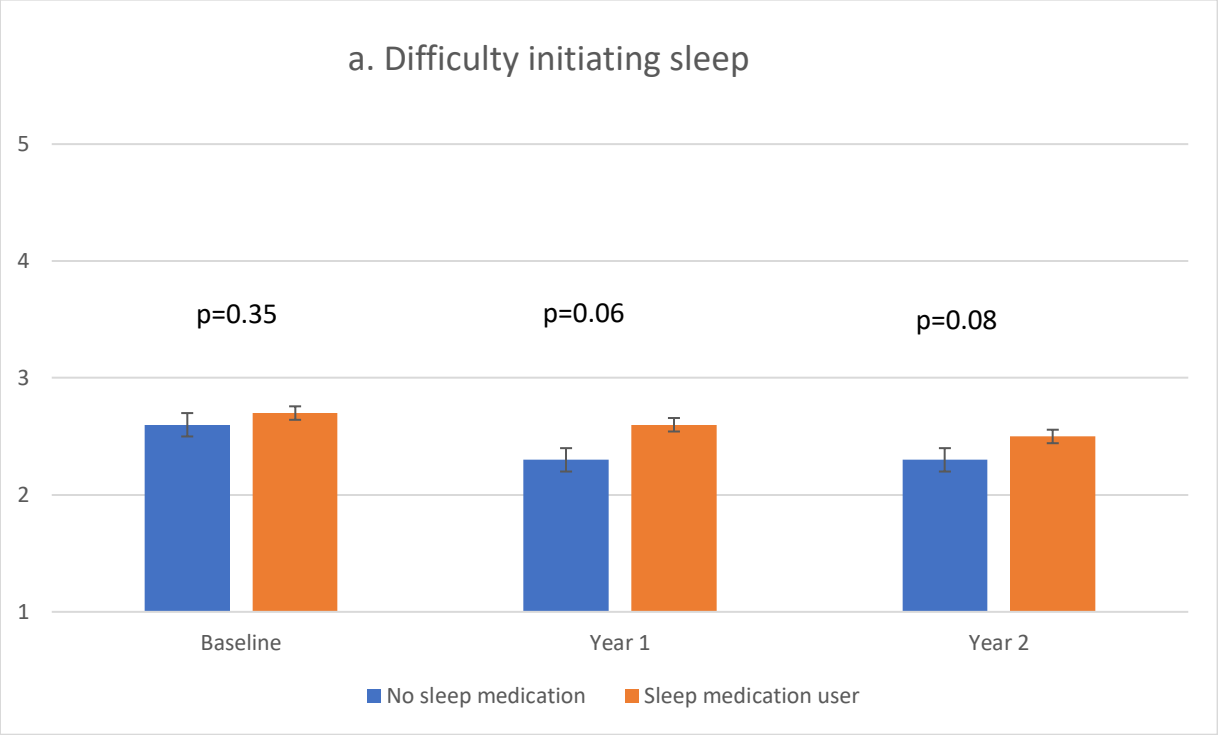
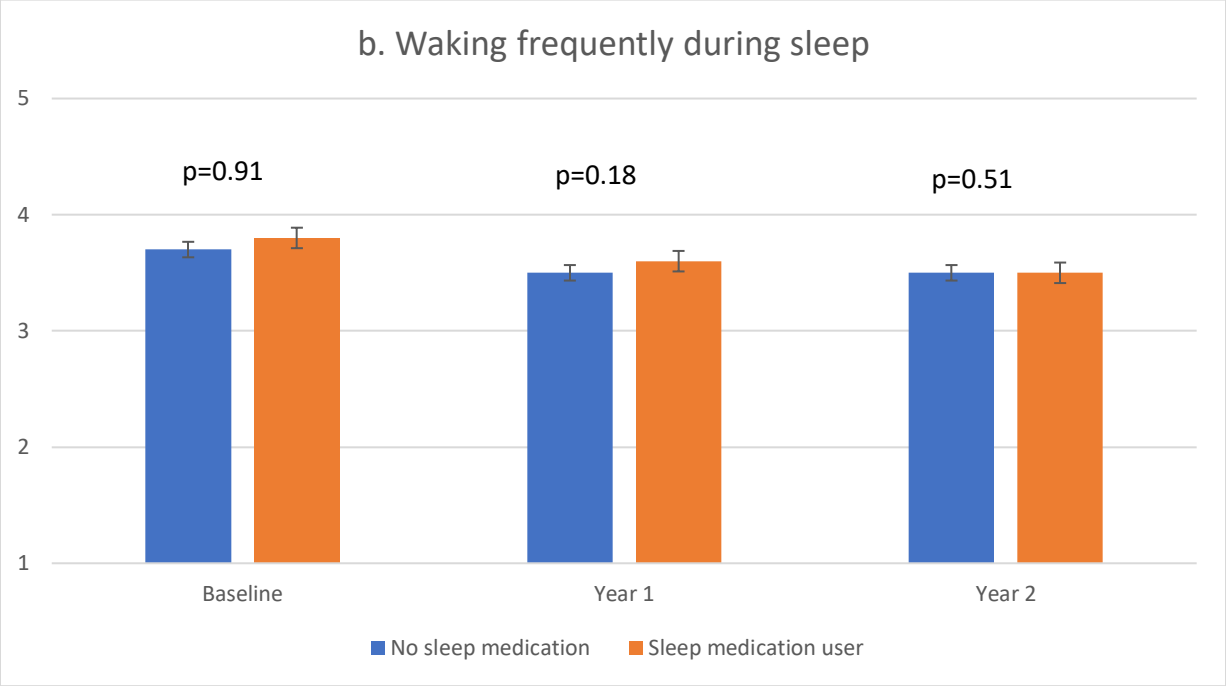


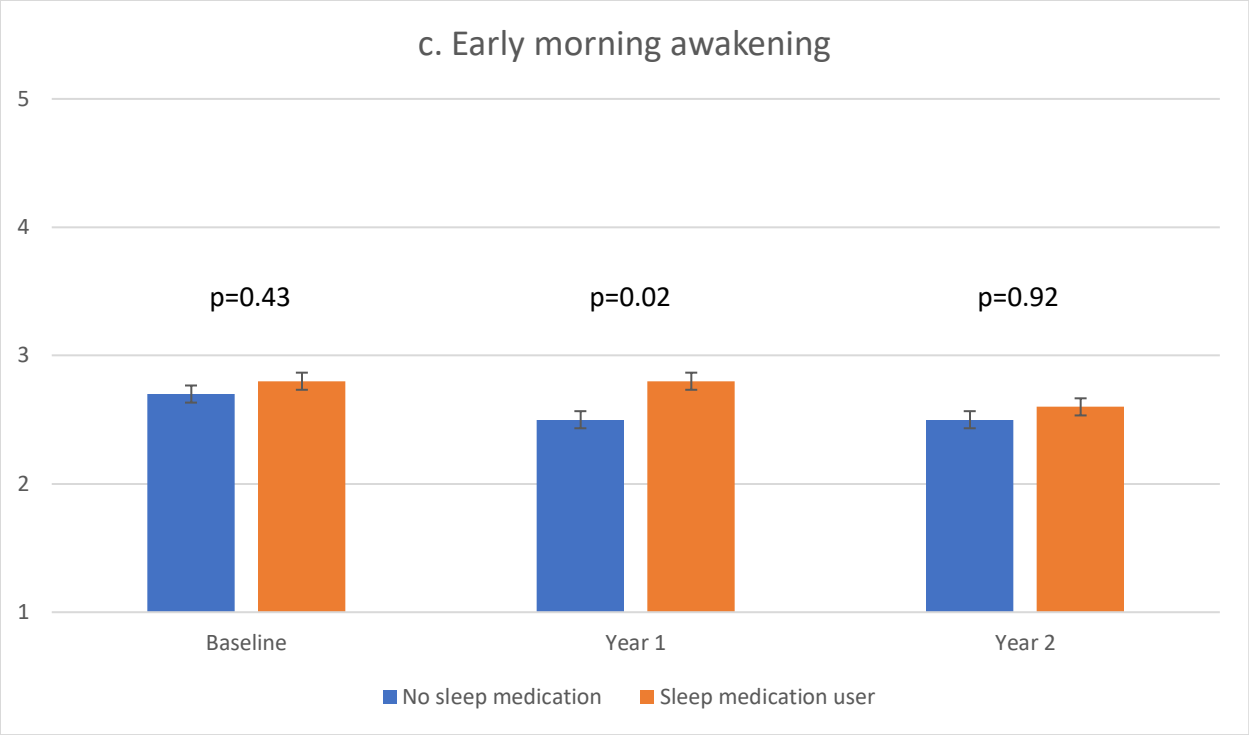
Figure 2 – Sleep Disturbance Ratings by Medication Exposure During Follow-up
Legend: Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week. Error bars represent standard errors. P-values at baseline, year 1 and year 2 comparing sleep medication users with non-users were estimated from the Wilcoxon Rank Sum test.



P-values for the differences between medication users and non-users for the change between baseline and one year = 0.19; baseline and two year = 0.55; and one year and two year = 0.73.



P-values for the differences between medication users and non-users for the change between baseline and one year = 0.41; baseline and two year = 0.98; and one year and two year = 0.55.



P-values for the differences between medication users and non-users for the change between baseline and one year = 0.13; baseline and two year = 0.46; * one year and two year = 0.03 (favoring non-use).

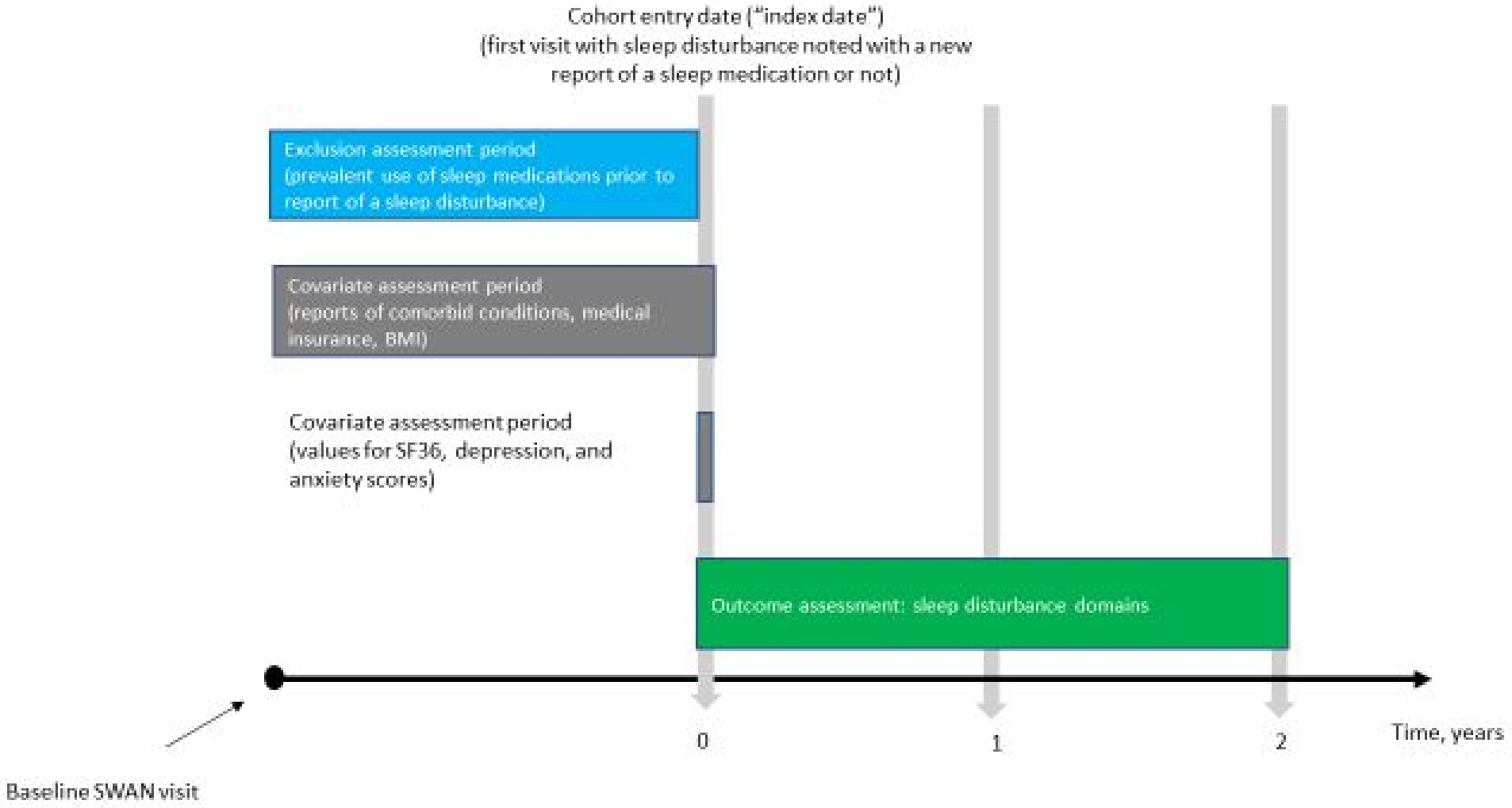
SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1: Design of the Target Trial and the Observational Corollary

Protocol component	Target Trial	Observational Corollary*
Study question	Are sleep medications effective over one year?	Same
Eligible criteria	Adult men and women reporting a sleep disturbance	Women in the SWAN cohort reporting a sleep disturbance
Other selection criteria	No use of sleep medications at baseline (or a sufficient washout period); no obstructive sleep apnea	No use of sleep medications at entry into SWAN; one-year follow-up data
Treatment strategies	Specific medication for sleep at a known effective dosage versus placebo	All known sleep medications at a variety of dosages versus no use of a sleep medication
Treatment assignment procedures	Randomization	Based on clinical evaluation during routine medical visits
Outcome	Sleep disturbance, self-reported and measured; assessed monthly	Self-reported sleep disturbances assessed one- and two-years after baseline
Balancing method	Randomization	Propensity score matching
Causal contrasts of interest	Intention to treat	Same

*Current study. SWAN, Study of Women Across the Nation.
Based on Hernan and Robins, Am J Epidemiology, 2016;183:758.

Supplementary Figure 1: Study Design



Supplementary Table 2: Baseline Demographics of Women in SWAN Who Were Included in the Current Analyses and Women Who Were Not

	Women Included in Study Cohort N=685	Women not Included in Study Cohort n=1846	SMD
	N (%) unless noted		
Age, mean (SD)	46.5 (2.7)	46.3 (2.7)	0.02
BMI, mean (SD)	28.2 (7.4)	28.5 (7.4)	0.03
Educational attainment			0.001
High school or less	141 (20.6)	443 (24.2)	
> high school	542 (79.4)	1387 (75.8)	
Ethnicity/race			0.001
African American	158 (23.1)	564 (30.6)	
White	394 (57.5)	867 (47.0)	
Chinese	45 (6.6)	145 (7.9)	
Hispanic	25 (3.7)	127 (6.9)	
Japanese	63 (9.2)	143 (7.8)	
Medical insurance	657 (95.9)	1696 (92.0)	0.06
Tobacco use			0.16
Never	347 (50.7)	1088 (59.0)	
Past/Current	337 (49.3)	756 (41.0)	
Alcohol use			0.16
None	294 (44.8)	877 (50.2)	
< 1 drink/week	60 (9.1)	179 (10.3)	
1-7 drinks/week	175 (26.6)	469 (28.9)	
>7 drinks/week	128 (19.5)	221 (12.7)	
Depression (CES-D), mean (SD)	12.3 (10.2)	10.7 (9.6)	0.31
Anxiety score, mean (SD)	3.1 (2.7)	2.5 (2.3)	0.30
Body pain, mean (SD)	65.3 (21.5)	68.8 (22.6)	0.27
Menopausal Status			0.05
Unknown	2 (0.3)	3 (0.2)	
Pre-menopausal	315 (46.3)	1023 (55.8)	
Early Peri-menopausal	364 (53.5)	808 (44.1)	
Diabetes	33 (4.8)	90 (4.9)	0.01
Hypertension	160 (23.5)	423 (23.1)	0.07
Osteoarthritis	150 (22.1)	312 (17.1)	0.02

Abbreviations: SMD, standardized mean difference; CES-D, Center for Epidemiologic Studies Depression Scale; BMI, Body Mass Index; SF36 Mental, Mental Component Score; and SF36 Physical, Physical Component Score.

Supplementary Table 3: Change in severity of sleep disturbances from baseline to year 1 for those propensity score matched at baseline minus 1 year

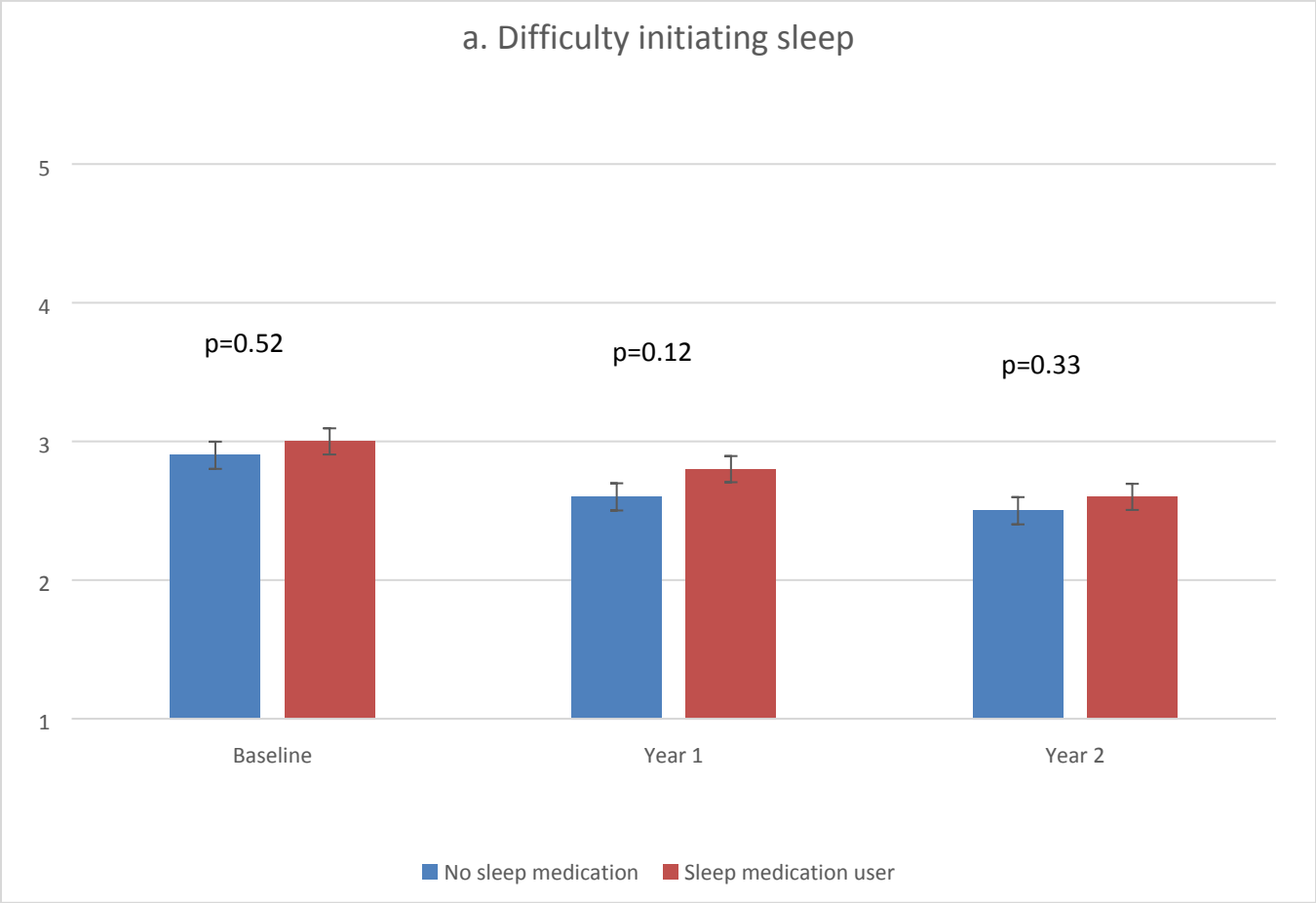
	Index Visit No Meds n=477		Index Visit Med Users n=253		Visit 1 year after No Meds n=477		Visit 1 year after Med Users n=253		Visit 2 year after No Meds n=361		Visit 2 year after Med Users n=197		P-value
	n	%	n	%	n	%	n	%	n	%	n	%	
Difficulty initiating sleep													0.17[1] 0.11[2] 0.83[3]
1	188	39.4%	81	32.0%	199	41.7%	96	37.9%	150	41.6%	71	36.0%	
2	87	18.2%	31	12.3%	90	18.9%	31	12.3%	74	20.5%	34	17.3%	
3	98	20.5%	44	17.4%	90	18.9%	41	16.2%	72	19.9%	35	17.8%	
4	57	11.9%	42	16.6%	47	9.9%	38	15.0%	28	7.8%	18	9.1%	
5	47	9.9%	55	21.7%	51	10.7%	47	18.6%	37	10.2%	39	19.8%	
Waking frequently during sleep													0.55[1] 0.14[2] 0.31[3]
1	58	12.2%	20	7.9%	78	16.4%	35	13.8%	60	16.6%	25	12.7%	
2	66	13.8%	23	9.1%	67	14.0%	27	10.7%	49	13.6%	25	12.7%	
3	102	21.4%	37	14.6%	96	20.1%	38	15.0%	75	20.8%	38	19.3%	
4	97	20.3%	73	28.9%	95	19.9%	52	20.6%	65	18.0%	41	20.8%	
5	154	32.3%	100	39.5%	141	29.6%	101	39.9%	112	31.0%	68	34.5%	
Early morning awakening													0.82[1] 0.02[2] 0.02[3]
1	161	33.8%	69	27.3%	192	40.3%	77	30.4%	135	37.4%	73	37.1%	
2	105	22.0%	37	14.6%	94	19.7%	50	19.8%	75	20.8%	32	16.2%	
3	100	21.0%	53	20.9%	77	16.1%	37	14.6%	69	19.1%	37	18.8%	
4	66	13.8%	45	17.8%	62	13.0%	34	13.4%	39	10.8%	21	10.7%	
5	45	9.4%	49	19.4%	52	10.9%	55	21.7%	43	11.9%	34	17.3%	
Any Complaint of 3 or more times week													0.10[1] 0.04[2]
Yes	279	58.5%	198	78.3%	264	55.3%	172	68.0%	197	54.6%	128	65.0%	0.53[3]

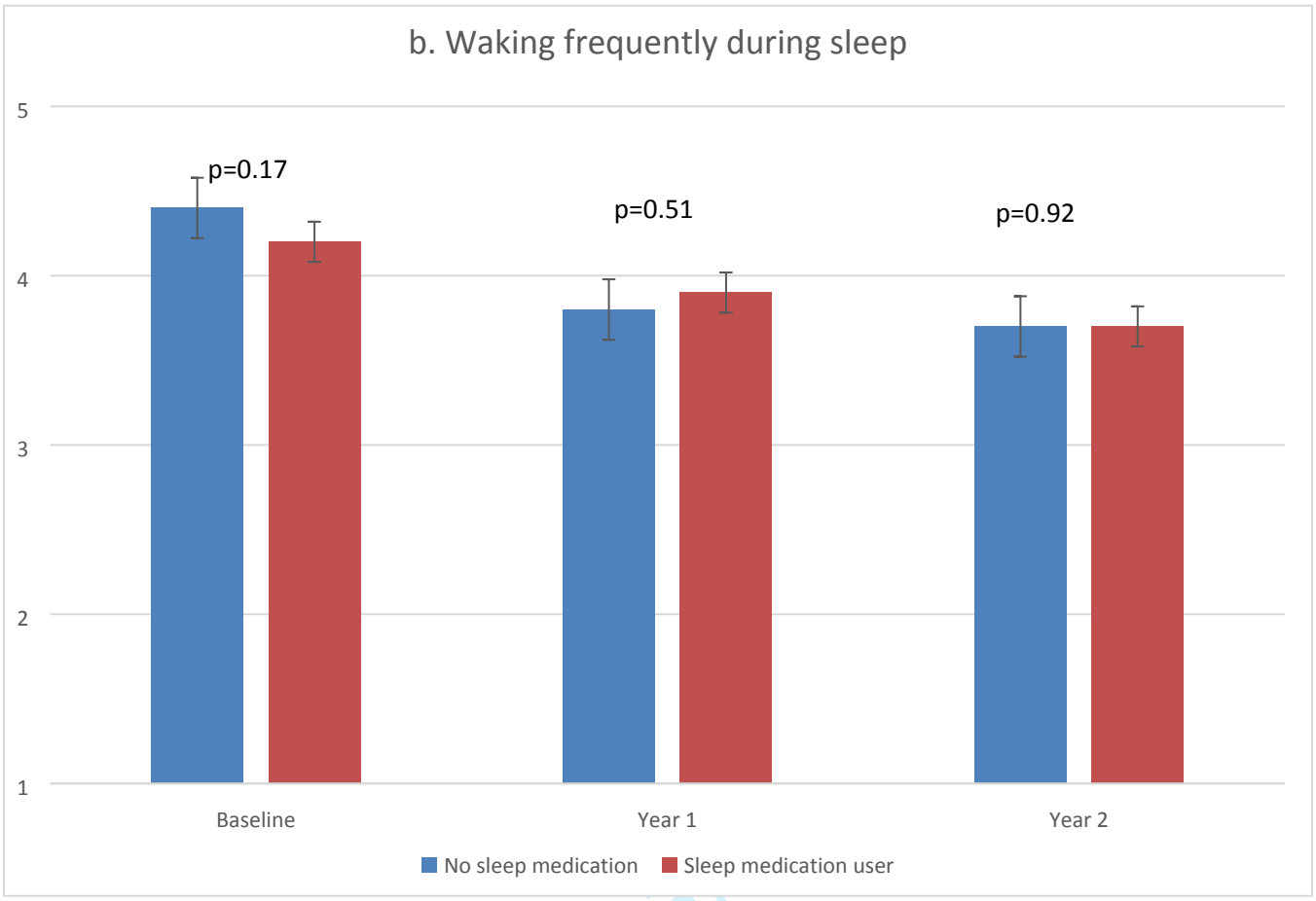
5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less <1 night/week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

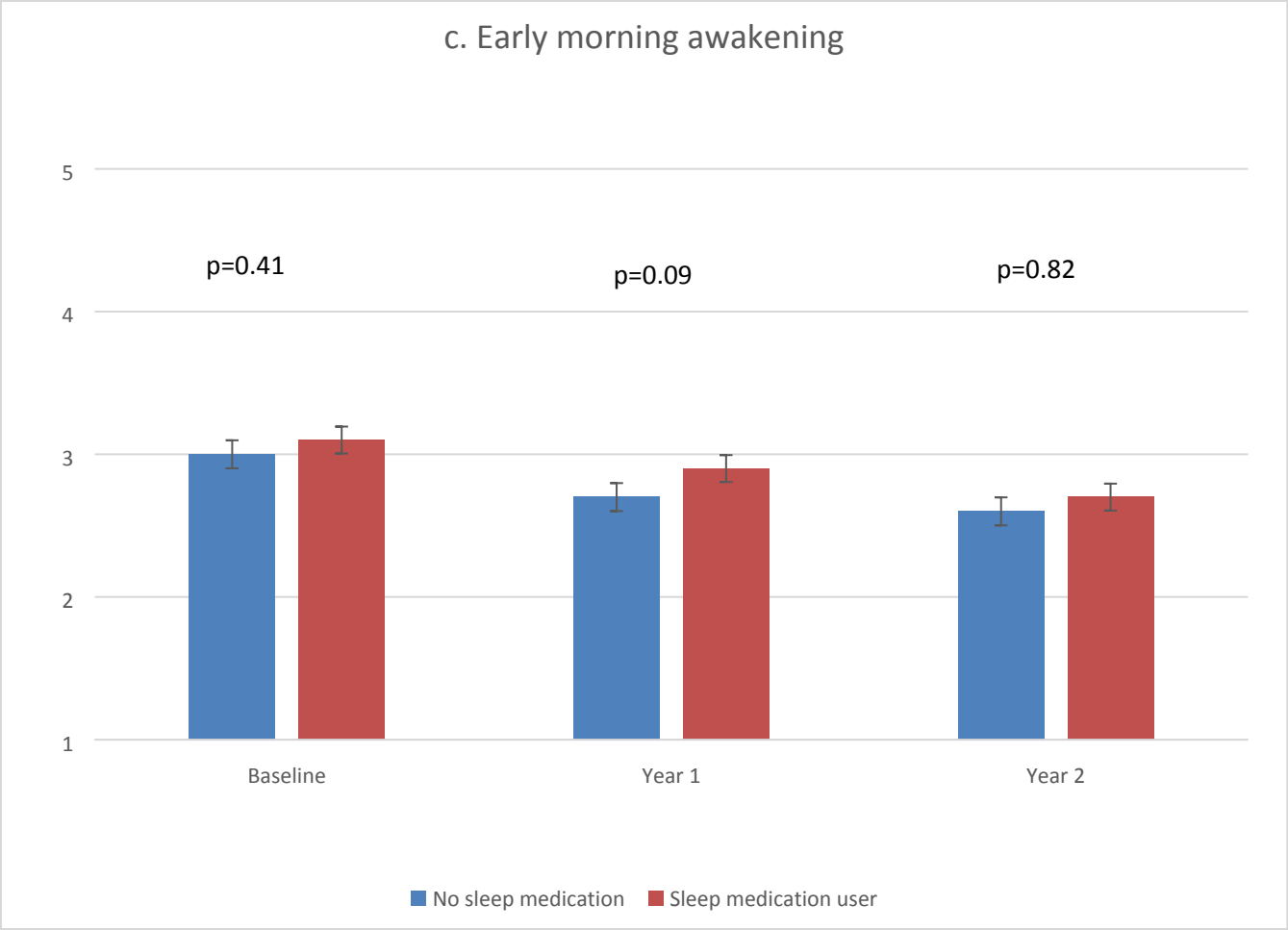
*p-values reflect change in severity of disturbances: [1]=baseline vs visit 1, [2]= baseline vs visit 2, [3]= visit 1 vs. visit 2.

Supplementary Figure 2. Sleep Disturbance Ratings by Medication Exposure During Follow-up In women who reported a 4 or 5 on any severity scale

Legend: Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on <1 night/week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week . Error bars represent standard errors. P-values estimated from the Wilcoxon Rank Sum test.






c. Early morning awakening
5
4
3
2
1
p=0.41
p=0.09
p=0.82
Baseline
Year 1
Year 2
No sleep medication
Sleep medication user

Supplementary Table 4: Change in severity of sleep disturbances from baseline to year 1 using a proportional odds analysis

	Year 0 versus Year 1	Year 1 versus Year 2	Year 0 versus Year 2
	Odds ratio (95% CI) comparing sleep medication users with non-users		
Difficulty initiating sleep	0.89 (0.58 – 1.30)	0.72 (0.42 – 1.30)	0.80 (0.61 – 1.00)
Waking frequently	0.92 (0.58 – 1.40)	1.44 (0.82 – 2.50)	1.20 (0.90 – 1.50)
Early morning awakening	0.75 (0.49 – 1.20)	1.20 (0.67 – 2.00)	0.94 (0.72 – 1.20)
Notes: The odds ratios represent the odds of a one level increase in the Likert scale, comparing sleep medication users to non-users.			

BMJ Open

Prescription Medications for Sleep Disturbances Among Midlife Women During Two Years of Follow-up: A SWAN Retrospective Cohort Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045074.R1
Article Type:	Original research
Date Submitted by the Author:	06-Jan-2021
Complete List of Authors:	Solomon, Daniel; Brigham and Women's Hospital, Division of Rheumatology, Immunology and Allergy; Brigham and Women's Hospital, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine Ruppert, Kristine; University of Pittsburgh Habel, Laurel; Kaiser Permanente Northern California, Division of Research Finkelstein, Joel; Massachusetts General Hospital Lian, Pam; University of Pittsburgh Joffe, Hadine; Harvard Univerisity, Psychiatry Kravitz, Howard M.; Rush Medical College of Rush University Department of Anesthesiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology
Keywords:	EPIDEMIOLOGY, SLEEP MEDICINE, THERAPEUTICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prescription Medications for Sleep Disturbances Among Midlife Women During Two Years of Follow-up: A SWAN Retrospective Cohort Study

Daniel H. Solomon (1), Kristine Ruppert (2), Laurel Habel (4), Joel Finkelstein (5), Pam Lian (3), Hadine Joffe (2), Howard M. Kravitz (6)

From: (1) Division of Rheumatology and Pharmacoepidemiology, Department of Medicine, Brigham and Women's Hospital, Boston MA; (2) Department of Epidemiology, University of Pittsburgh, Pittsburgh PA; (3) Division of Research, Kaiser Permanente Northern California, Oakland CA; (4) Endocrine Unit, Massachusetts General Hospital, Boston MA; (5) Department of Psychiatry, and Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital, Boston MA; (6) Department of Psychiatry and Department of Preventive Medicine, Rush University Medical Center, Chicago, IL

Support: 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 Aging (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). 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. Clinical Centers: University of Michigan, Ann Arbor – Siobán Harlow, PI 2011 – present, MaryFran Sowers, PI 1994-2011; Massachusetts General Hospital, Boston, MA – Joel Finkelstein, PI 1999 – present; Robert Neer, PI 1994 – 1999; Rush University, Rush University Medical Center, Chicago, IL – 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, NY – 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, PA – Karen Matthews, PI. NIH Program Office: National Institute on Aging, Bethesda, MD – Chhanda Dutta 2016- present; Winifred Rossi 2012–2016; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD – Program Officers. Central Laboratory: University of Michigan, Ann Arbor – Daniel McConnell (Central Ligand Assay Satellite Services). Coordinating Center: University of Pittsburgh, Pittsburgh, PA – Maria Mori Brooks, PI 2012 - present; Kim Sutton-Tyrrell, PI 2001 – 2012; New England Research Institutes, Watertown, MA - Sonja McKinlay, PI 1995 – 2001. Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair

Potential Conflicts: Dr. Solomon 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.

Correspondence: Daniel H. Solomon, Division of Rheumatology, Brigham and Women's Hospital, 60 Fenwood Road, Boston MA, 02115. T: 617-732-5356. F: 617-732-5505. E: dsolomon@bwh.harvard.edu

Words: 3522; Tables: 4, Figures: 2; Citations: 31

Supplements: Yes

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 US.

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 two weeks – difficulty initiating asleep, 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 one-year (primary outcome) and two-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 asleep of 2.7 (SD 1.5), waking frequently 3.8 (SD 1.3), and early morning awakening 2.8 (SD 1.5); non-users ratings were 2.6 (SD 1.5), 3.7 (SD 1.3), and 2.7 (SD 1.4), respectively. After one year, ratings for medication users were 2.6 (SD 1.6) for initiating asleep, 3.6 (SD 1.5) for waking frequently, and 2.8 (SD 1.5) for early morning awakening; for non-users, the mean ratings were 2.3 (SD 1.4), 3.5 (SD 1.4), and 2.5 (SD 1.5), respectively. None of the one-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 one and two years. The effectiveness of long-term sleep medication use should be re-examined.

Article summary: 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 United States 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 pharmacologic treatment options have been developed for sleep.

Prescription sleep medications consist of benzodiazepines (BZDs), Z-drugs (selective benzodiazepine receptor agonists that include zolpidem, zaleplon, and eszopiclone) and other agents mostly used off-label to promote sleep through a variety of other mechanisms. Randomized controlled trials 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-randomized designs to test the benefits of drugs.[10] We assessed the potential benefits of sleep medications among a large and diverse cohort of mid-life 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 Hernan and Robins.[11] In this study paradigm, a target randomized controlled trial 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 **Supplementary 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. Further, we described the study design using standardized illustrations as suggested by Schneeweiss and colleagues (see **Supplementary 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 multicenter, multi-ethnic/multi-racial 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 to 52 years; intact uterus and at least one ovary, not using sex steroid hormones or pregnant, breastfeeding or lactating at enrollment or within the previous three 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 3 nights per week during a two-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 benzodiazepine receptor agonists, and other hypnotics. The full list of medications considered included the following BZDs: estazolam, flurazepam, lorazepam, temazepam, and triazolam; , selective benzodiazepine 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. Further, 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 two weeks. They used a five-point Likert scale to report on each type of disturbance, where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.[17-19] We considered the results at one-year to be the primary outcome and two-years to be the secondary outcome. For the two-year outcome, only women who had both year one and year two results were analyzed.

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 CES-D,[20] anxiety with the GAD-7,[21] and the SF-36 scales were used to measure pain, mental function and physical function.[22]

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.[23] A propensity score estimates the likelihood that women would start a sleep medication, with values ranging from zero to one. 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.[24] We attempted to match 2 non-users for each user using a "greedy matching" algorithm, with a maximum caliper of 0.2 of a standard deviation of the logit of the propensity score.[25]

After matching, we examined baseline characteristics for balance using standardized 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 one- and two-years, estimating means and standard deviations, and the changes in sleep disturbance from baseline to one year and one year to two years. These changes were estimated and compared across medication exposure groups, using a mixed model regression. 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 percent 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. As well, we 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 4 or 5 on at least one sleep domain. This definition is consistent with the frequency criterion for clinically significant sleep difficulty (e.g., insomnia disorder).[26,27] We compared no medication use to specific sleep medications, benzodiazepines and selective benzodiazepine receptor agonists. Finally, we ran models adjusted for SWAN site and estrogen replacement therapy. Such analyses retained the propensity score match.

All analyses were conducted using SAS (Cary NC). All p-values were nominal and not adjusted for multiple comparisons, as these were post-hoc exploratory analyses.

RESULTS

We identified 2,531 potentially eligible women in SWAN who reported the severity of a sleep disturbance at some point during the twenty-one years of follow-up, 1995-2016 (see **Figure 1**). We applied the exclusion criteria and found 1,528 women who were analyzed 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 Supplementary Table 2). Thus, we propensity matched the 238,

attempting, attempting to find 2 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 1,846 potentially eligible women not included in the analysis (see **Supplementary Table 3**). 100% 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 standardized mean differences were $\leq$ 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/m2 (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 peri-menopause. 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 one 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 group (medication users or non-users) were 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, SD 1.5) compared to those who did (mean 2.8, SD 1.5; p = 0.02). The secondary two-year outcomes were similar to the one-year results; none demonstrated statistically significant reductions in sleep disturbances among sleep medication users.

Several secondary analyses were pursued. First, we examined the distribution of Likert scores at baseline and one 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 benzodiazepine receptor agonists and other hypnotics (see **Table 4**); no differences were observed in the change from baseline to one 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 **Supplementary 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 Supplementary 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 **Supplementary Figure 2**). Finally, analyses that also included site and estrogen use gave similar results (see **Supplementary Table 6**).

DISCUSSION

Sleep difficulties are common.[1,28] 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.[29] Most data regarding their efficacy derive from short term studies (i.e., 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-characterized middle-aged community-dwelling women with sleep 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 one year later. While there are good data from randomized controlled trials 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 twenty percent 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 randomized controlled trial (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 analyzed appropriately to complement randomized trials.[10] The FDA has published a framework for generating evidence from real-world observational data sets,[30] with the hope that such analyses will allow clinicians to better understand the benefits and risks of drugs in typical practice.

We used rigorous epidemiologic methods and analyzed a well characterized cohort of women, but as with all observational studies there are limitations to recognize. The use of sleep medications was not randomized. 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 pre-defined 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 one-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 two-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 two-year analysis.

Sleep disturbances were self-reported, without any objective measures of sleep. This may have introduced mis-classification, 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 analyzed as a continuous variable in the mixed regression models, however analyses that retained the five categories gave similar negative results (see Table 3 and Supplementary 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, i.e., at 8 or 12 weeks. Women only reported medication use and sleep disturbances at annual visits and thus interim outcomes (i.e., at six 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 characterized cohort of women during a high-risk period for sleep disturbance. It is known that women going through the midlife often note sleep disturbances.[31] As well, we studied women of several races and ethnicities, enhancing the generalizability 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 randomized controlled clinical trials. The current observational study does not support use of sleep medications over the long-term, as there were no self-reported differences at one- or two-years of follow-up comparing sleep medication users to non-users. While we used rigorous epidemiologic methods, the findings reported herein are based on a non-randomized 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 behavioral therapy.[32] 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 one- and two-years should help inform clinicians and patients considering initiating pharmacologic treatment for midlife women who have sleep complaints.

Legends:

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.

Figure 2. These 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 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week. Error bars represent standard errors. 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 one year = 0.19; baseline and two year = 0.55; and one year and two year = 0.73. In Panel B, P-values for the differences between medication users and non-users for the change between baseline and one year = 0.41; baseline and two year = 0.98; and one year and two year = 0.55. In Panel C, P-values for the differences between medication users and non-users for the change between baseline and one year = 0.13; baseline and two year = 0.46; * one year and two year = 0.03 (favoring non-use).

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	SMD
		N (%) unless noted		
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
SF36-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
Pre-menopausal	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
Post-menopausal	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

Abbreviations: SMD, standardized mean difference; CES-D, Center for Epidemiologic Studies Depression Scale; BMI, Body Mass Index; SF36 Mental, Mental Component Score; and SF36 Physical, Physical Component Score. 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 nonsteroidal anti-inflammatory drugs. The CES-D is a 20-item scale with a range of 0 to 60 (REF 20). The anxiety score (GAD-7) is a 7-item scale with a range of 0-21 (REF 21). The SF-36 bodily pain score includes two items with a range of 0 to 100; SF-36 mental component score is a 5-item scale with a range of 0 to 100; and SF-36 physical function is a 10-item scale with a range of 0 to 100 (REF 21).

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 (SD)*	2.6 (1.5)	2.7 (1.5)	0.08
Waking frequently, mean (SD)*	3.7 (1.3)	3.8 (1.3)	0.03
Early morning awakening, mean (SD)*	2.7 (1.4)	2.8 (1.5)	0.07
Trouble initiating sleep, at least 3 nights per week, n (%)	137 (30.7)	82 (34.5)	0.07
Waking frequently, at least 3 nights per week, n (%)	291 (65.1)	158 (66.4)	0.008
Early morning awakening, at least 3 nights per week, n (%)	135 (30.2)	81 (34.0)	0.07
Any disturbance, at least 3 nights per week, n (%)	322 (72.0)	183 (76.9)	0.08

Abbreviations: SD, standard deviation; SMD, standardized mean difference. *Mean calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

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	
Sleep Disturbance	N	%	n	%	n	%	n	%	n	%	n	%
Difficulty initiating sleep (per week)												
1 (no difficulty)	154	34.5%	81	34.0%	190	42.5%	94	39.5%	156	44.2%	70	37.4%
2 (= <1 night/week)	74	16.6%	31	13.0%	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 (3-4nights/week)	74	16.6%	39	16.4%	49	11.0%	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	16.0%	78	22.1%	36	19.3%
4 (3-4 nights/week)	118	26.4%	69	29.0%	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	32.0%	65	34.8%
Early morning awakening												
1 (no difficulty)	127	28.4%	69	29.0%	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	16.0%
3 (1-2 nights/week)	102	22.8%	51	21.4%	67	15.0%	35	14.7%	67	19.0%	34	18.2%
4 (3-4 nights/week)	76	17.0%	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	72.0%	183	76.9%	273	61.1%	159	66.8%	203	57.5%	122	65.2%

Abbreviations: SD, standard deviation. Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less than 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

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	BZD Users	No Sleep Medications	BZD Users	P-value*
	n=447	n=87	n=447	n=87	
Difficulty initiating sleep, mean (SD)	2.6 (1.5)	2.2 (1.6)	2.3 (1.4)	2.6 (1.6)	0.71
Waking frequently during sleep, mean (SD)	3.7 (1.3)	3.8 (1.3)	3.5 (1.4)	3.3 (1.4)	0.24
Early morning awakening, mean (SD)	2.7 (1.4)	2.6 (1.5)	2.5 (1.5)	2.6 (1.6)	0.17
	No Sleep Medications	Z-drugs + other hypnotics	No Sleep Medications	Z-drugs + other hypnotics	
	n=447	n=151	n=447	n=151	
Difficulty initiating sleep, mean (SD)	2.6 (1.5)	2.7 (1.5)	2.3 (1.4)	2.6 (1.6)	0.12
Waking frequently during sleep, mean (SD)	3.7 (1.3)	3.8 (1.2)	3.5 (1.4)	3.8 (1.4)	0.05
Early morning awakening, mean (SD)	2.7 (1.4)	2.9 (1.5)	2.5 (1.5)	2.8 (1.5)	0.28

Abbreviations: SD, standard deviation; BZD, benzodiazepine; Z-drugs (selective benzodiazepine 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 1 night per week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 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 one.

Contributorship statement:

Daniel H. Solomon: Design, analysis, drafting and revising manuscript.
Kristine Ruppert: Analysis and revising manuscript.
Laurel Habel: Design, analysis and revising manuscript.
Joel Finkelstein: Data collection, design, and revising manuscript.
Pam Lian: Analysis and revising manuscript.
Hadine Joffe: Design and revising manuscript.
Howard M. Kravitz: Data collection, design and revising manuscript.

Competing interests: Dr. Solomon 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.

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 Aging (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). 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.

Data sharing statement: Data from SWAN are available for qualified researchers.

Ethics statement: 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 Center – 13021201-IRB01-AM04; University of California, Davis – 260339-17; UCLA – 11-002274-AM-00009.

REFERENCES

1. Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. *Sleep.* 2014;37(2):343-349.
2. Kaufmann CN, Spira AP, Alexander GC, Rutkow L, Mojtabai R. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. *Pharmacoepidemiology and drug safety.* 2016;25(6):637-645.
3. Ford ES, Wheaton AG, Cunningham TJ, Giles WH, Chapman DP, Croft JB. 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(8):1283-1293.
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(4):780-784.
5. Schroeck JL, Ford J, Conway EL, et al. Review of Safety and Efficacy of Sleep Medicines in Older Adults. *Clin Ther.* 2016;38(11):2340-2372.
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(7):793-799.
7. Randall S, Roehrs TA, Roth T. Efficacy of eight months of nightly zolpidem: a prospective placebo-controlled study. *Sleep.* 2012;35(11):1551-1557.
8. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. 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(2):307-349.
9. Jaussent I, Ancelin ML, 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, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clinical pharmacology and therapeutics.* 2019;105(4):867-877.
11. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764.
12. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. *Ann Intern Med.* 2019;170(6):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(25):2366-2373.
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(5):999-1015.
15. Sowers M, Crawford S, Sternfeld B, et al. *Design, survey, sampling and recruitment methods of SWAN: a multi-center, multi-ethnic, community based cohort study of women and the menopausal transitioin. 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. *Journal of Clinical Pharmacy and Therapeutics.* 2008;2:227-237.
17. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol.* 1988;41(4):313-321.

18. Levine DW, Dailey ME, Rockhill B, Tipping D, Naughton MJ, Shumaker SA. Validation of the Women's Health Initiative Insomnia Rating Scale in a multicenter controlled clinical trial. *Psychosom Med.* 2005;67(1):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(2):137-148.
20. Carleton RN, Thibodeau MA, Teale MJ, 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(3):e58067.
21. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092-1097.
22. Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976).* 2000;25(24):3130-3139.
23. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *American Journal of Epidemiology.* 2006;163(12):1149-1156.
24. Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biometrical journal.* 2009;51(1):171-184.
25. Parsons LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques. SAS. www2.sas.com/proceedings/sugi26/p214-26.pdf. Accessed 2014.
26. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. 5th ed (DSM-5).* Arlington, VA2013.
27. American Academy of Sleep Medicine. *International Classification of Sleep Disorders. 3rd ed.* Darien, IL: American Academy of Sleep Medicine; 2014.
28. Kaufman DW, Kelly JP, Rosenberg L, Anderson TE, Mitchell AA. Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey.[see comment]. *JAMA.* 2002;287(3):337-344.
29. Verster JC, Mooren L, Bervoets AC, Roth T. 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(3):e12622.
30. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. In:2018.
31. 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(3):235-250.
32. Medalie L, Cifu AS. Management of Chronic Insomnia Disorder in Adults. *JAMA.* 2017;317(7):762-763.

Figure 1

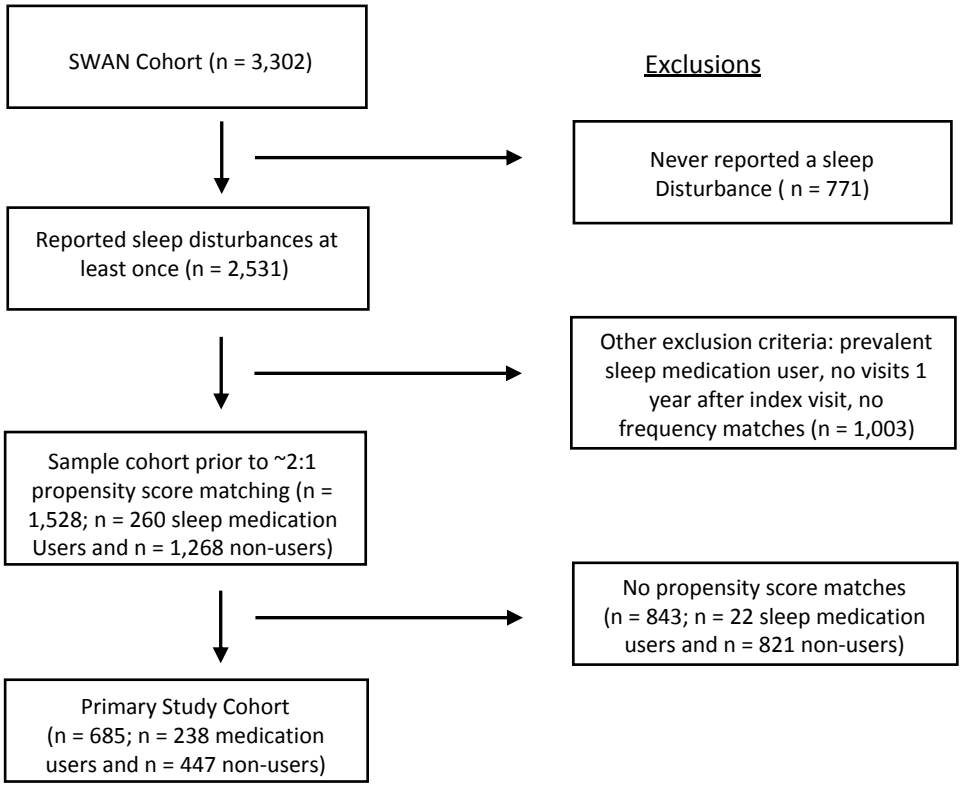
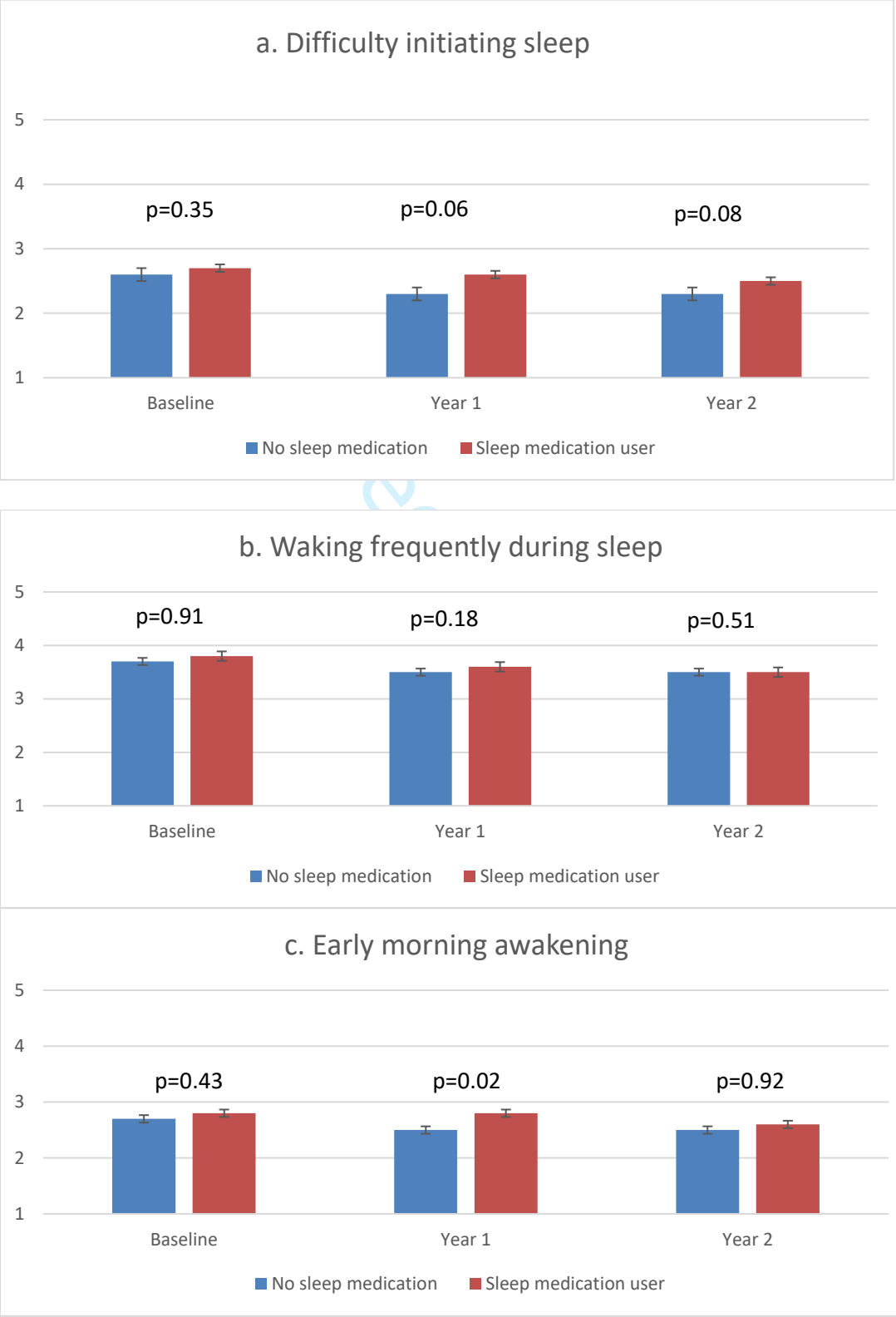

SWAN Cohort (n = 3,302)
Exclusions
Never reported a sleep Disturbance (n = 771)
Reported sleep disturbances at least once (n = 2,531)
Other exclusion criteria: prevalent sleep medication user, no visits 1 year after index visit, no frequency matches (n = 1,003)
Sample cohort prior to ~2:1 propensity score matching (n = 1,528; n = 260 sleep medication Users and n = 1,268 non-users)
No propensity score matches (n = 843; n = 22 sleep medication users and n = 821 non-users)
Primary Study Cohort (n = 685; n = 238 medication users and n = 447 non-users)

Figure 2


a. Difficulty initiating sleep
5
4
3
2
1
p=0.35
p=0.06
p=0.08
Baseline
Year 1
Year 2
No sleep medication
Sleep medication user
b. Waking frequently during sleep
5
4
3
2
1
p=0.91
p=0.18
p=0.51
Baseline
Year 1
Year 2
No sleep medication
Sleep medication user
c. Early morning awakening
5
4
3
2
1
p=0.43
p=0.02
p=0.92
Baseline
Year 1
Year 2
No sleep medication
Sleep medication user

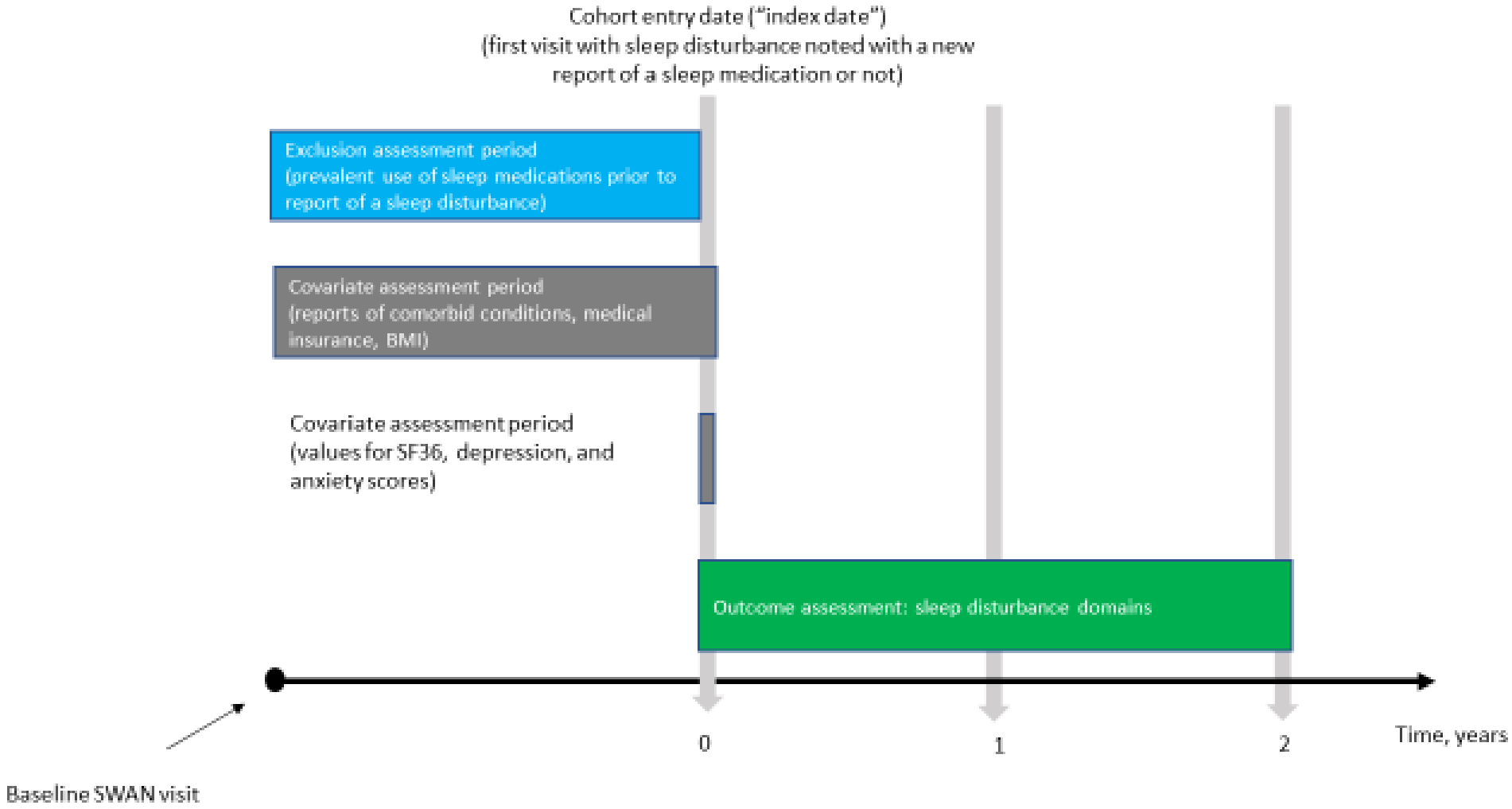
SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1: Design of the Target Trial and the Observational Corollary

Protocol component	Target Trial	Observational Corollary*
Study question	Are sleep medications effective over one year?	Same
Eligible criteria	Adult men and women reporting a sleep disturbance	Women in the SWAN cohort reporting a sleep disturbance
Other selection criteria	No use of sleep medications at baseline (or a sufficient washout period); no obstructive sleep apnea	No use of sleep medications at entry into SWAN; one-year follow-up data
Treatment strategies	Specific medication for sleep at a known effective dosage versus placebo	All known sleep medications at a variety of dosages versus no use of a sleep medication
Treatment assignment procedures	Randomization	Based on clinical evaluation during routine medical visits
Outcome	Sleep disturbance, self-reported and measured; assessed monthly	Self-reported sleep disturbances assessed one- and two-years after baseline
Balancing method	Randomization	Propensity score matching
Causal contrasts of interest	Intention to treat	Same

*Current study. SWAN, Study of Women Across the Nation.
Based on Hernan and Robins, Am J Epidemiology, 2016;183:758.

Supplementary Figure 1: Study Design



Supplementary Table 2: Baseline Demographics of Women in SWAN Examined in the Primary Cohort

	Total N=1528	No Sleep Medication n=1268	Sleep Medication User n=260	SMD
		N (%) unless noted		
Age, mean (SD)	49.8 (8.4)	49.9 (8.5)	49.3 (7.6)	0.08
BMI, mean (SD)	28.8 (7.2)	28.7 (7.1)	29.2 (7.7)	0.07
Educational attainment				
High school or less	319 (20.9)	261 (20.6)	58 (22.3)	0.01
> high school	1201 (78.6)	999 (78.8)	202 (77.7)	
Ethnicity/race				
African American	406 (26.6)	345 (27.2)	61 (23.5)	0.09
White	782 (51.2)	634 (50.0)	148 (56.9)	0.14
Chinese	135 (8.8)	119 (9.4)	16 (6.2)	0.12
Hispanic	44 (2.9)	33 (2.6)	11 (4.2)	0.09
Japanese	161 (10.5)	137 (10.8)	24 (9.2)	0.05
Medical insurance	1438 (94.1)	1186 (93.5)	252 (96.9)	0.16
Marital status				
Single	194 (12.7)	155 (12.2)	39 (15.0)	0.06
Married	991 (64.9)	831 (65.6)	160 (61.5)	0.15
Separated	43 (2.8)	33 (2.6)	10 (3.9)	0.15
Widowed	67 (4.4)	54 (4.3)	13 (5.0)	0.08
Divorced	232 (15.2)	194 (15.3)	38 (14.6)	0.03
Tobacco use				
Never	895 (58.6)	761 (60.0)	113 (43.5)	0.17
Past/Current	629 (41.2)	504 (39.8)	125 (48.1)	
Alcohol use				0.05
None	733 (49.6)	621 (50.7)	112 (44.3)	0.13
<1 drink/week	373 (25.3)	321 (26.2)	52 (20.6)	0.13
1-7 drinks/week	252 (17.1)	190 (15.5)	62 (24.5)	0.23
>7 drinks/week	119 (8.1)	92 (7.5)	27 (10.7)	0.11
Depression (CES-D), mean (SD)	9.4 (9.4)	8.5 (8.6)	14.1 (11.4)	0.57
Anxiety score, mean (SD)	2.4 (2.4)	2.1 (2.2)	3.4 (2.8)	0.49
Body pain, mean (SD)	68.1 (22.4)	69.7 (21.7)	60.1 (24.0)	0.42
SF36-Mental, mean (SD)	49.1 (10.4)	49.9 (9.9)	45.0 (11.5)	0.45
SF36-Physical, mean (SD)	49.7 (9.9)	50.2 (9.5)	47.4 (11.5)	0.27
Menopausal Status				
Unknown	138 (9.0)	98 (7.7)	40 (15.4)	0.24
Pre-menopausal	113 (7.4)	102 (8.0)	11 (4.2)	0.16
Early/Late Peri-menopausal	597 (39.1)	508 (40.1)	89 (34.2)	0.12
Surgical menopause	54 (3.5)	42 (3.3)	12 (34.2)	0.07
Post-menopausal	625 (40.9)	517 (40.8)	108 (41.5)	0.02

Diabetes	152 (10.0)	123 (9.7)	29 (11.2)	0.05
Hypertension	616 (40.3)	488 (38.5)	128 (49.2)	0.22
Osteoarthritis	565 (37.0)	448 (35.3)	117 (45.0)	0.19
Cancer, current	26 (3.5)	12 (1.0)	14 (5.4)	0.16
Any antidepressant	28 (1.8)	11 (0.9)	17 (6.5)	0.30
Any analgesic	72 (4.7)	65 (5.1)	7 (2.7)	0.13

Supplementary Table 3: Baseline Demographics of Women in SWAN Who Were Included in the Current Analyses and Women Who Were Not

	Women Included in Study Cohort N=685	Women not Included in Study Cohort n=1846	SMD
	N (%) unless noted		
Age, mean (SD)	46.5 (2.7)	46.3 (2.7)	0.02
BMI, mean (SD)	28.2 (7.4)	28.5 (7.4)	0.03
Educational attainment			0.001
High school or less	141 (20.6)	443 (24.2)	
> high school	542 (79.4)	1387 (75.8)	
Ethnicity/race			0.001
African American	158 (23.1)	564 (30.6)	
White	394 (57.5)	867 (47.0)	
Chinese	45 (6.6)	145 (7.9)	
Hispanic	25 (3.7)	127 (6.9)	
Japanese	63 (9.2)	143 (7.8)	
Medical insurance	657 (95.9)	1696 (92.0)	0.06
Tobacco use			0.16
Never	347 (50.7)	1088 (59.0)	
Past/Current	337 (49.3)	756 (41.0)	
Alcohol use			0.16
None	294 (44.8)	877 (50.2)	
< 1 drink/week	60 (9.1)	179 (10.3)	
1-7 drinks/week	175 (26.6)	469 (28.9)	
>7 drinks/week	128 (19.5)	221 (12.7)	
Depression (CES-D), mean (SD)	12.3 (10.2)	10.7 (9.6)	0.31
Anxiety score, mean (SD)	3.1 (2.7)	2.5 (2.3)	0.30
Body pain, mean (SD)	65.3 (21.5)	68.8 (22.6)	0.27
Menopausal Status			0.05
Unknown	2 (0.3)	3 (0.2)	
Pre-menopausal	315 (46.3)	1023 (55.8)	
Early Peri-menopausal	364 (53.5)	808 (44.1)	
Diabetes	33 (4.8)	90 (4.9)	0.01
Hypertension	160 (23.5)	423 (23.1)	0.07
Osteoarthritis	150 (22.1)	312 (17.1)	0.02

Abbreviations: SMD, standardized mean difference; CES-D, Center for Epidemiologic Studies Depression Scale; BMI, Body Mass Index; SF36 Mental, Mental Component Score; and SF36 Physical, Physical Component Score.

Supplementary Table 4: Change in severity of sleep disturbances from baseline to year 1 for those propensity score matched at baseline minus 1 year

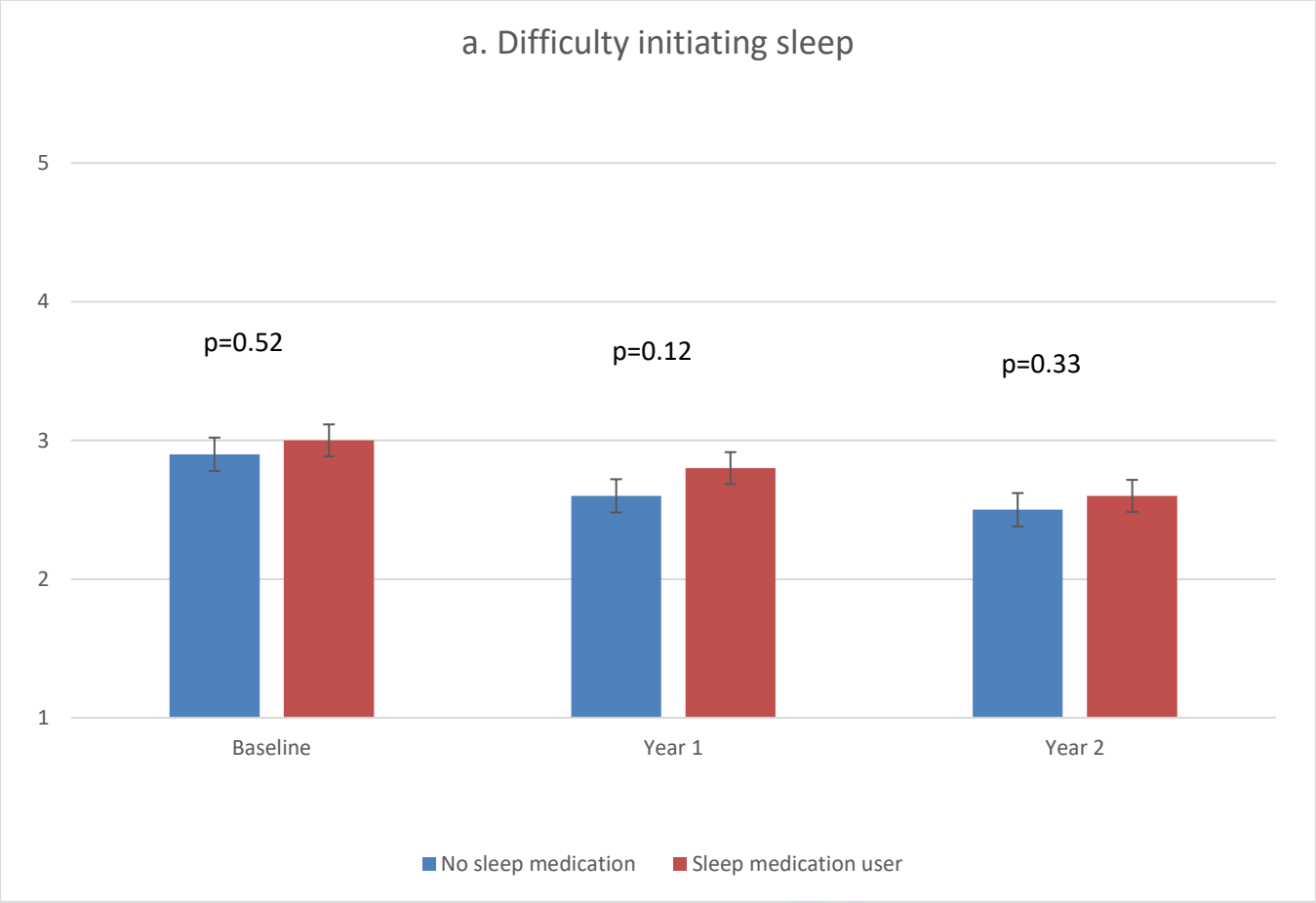
	Index Visit No Meds n=477		Index Visit Med Users n=253		Visit 1 year after No Meds n=477		Visit 1 year after Med Users n=253		Visit 2 year after No Meds n=361		Visit 2 year after Med Users n=197		P-value
	n	%	n	%	n	%	n	%	n	%	n	%	
Difficulty initiating sleep													0.17[1] 0.11[2] 0.83[3]
1	188	39.4%	81	32.0%	199	41.7%	96	37.9%	150	41.6%	71	36.0%	
2	87	18.2%	31	12.3%	90	18.9%	31	12.3%	74	20.5%	34	17.3%	
3	98	20.5%	44	17.4%	90	18.9%	41	16.2%	72	19.9%	35	17.8%	
4	57	11.9%	42	16.6%	47	9.9%	38	15.0%	28	7.8%	18	9.1%	
5	47	9.9%	55	21.7%	51	10.7%	47	18.6%	37	10.2%	39	19.8%	
Waking frequently during sleep													0.55[1] 0.14[2] 0.31[3]
1	58	12.2%	20	7.9%	78	16.4%	35	13.8%	60	16.6%	25	12.7%	
2	66	13.8%	23	9.1%	67	14.0%	27	10.7%	49	13.6%	25	12.7%	
3	102	21.4%	37	14.6%	96	20.1%	38	15.0%	75	20.8%	38	19.3%	
4	97	20.3%	73	28.9%	95	19.9%	52	20.6%	65	18.0%	41	20.8%	
5	154	32.3%	100	39.5%	141	29.6%	101	39.9%	112	31.0%	68	34.5%	
Early morning awakening													0.82[1] 0.02[2] 0.02[3]
1	161	33.8%	69	27.3%	192	40.3%	77	30.4%	135	37.4%	73	37.1%	
2	105	22.0%	37	14.6%	94	19.7%	50	19.8%	75	20.8%	32	16.2%	
3	100	21.0%	53	20.9%	77	16.1%	37	14.6%	69	19.1%	37	18.8%	
4	66	13.8%	45	17.8%	62	13.0%	34	13.4%	39	10.8%	21	10.7%	
5	45	9.4%	49	19.4%	52	10.9%	55	21.7%	43	11.9%	34	17.3%	
Any Complaint of 3 or more times week													0.10[1] 0.04[2]
Yes	279	58.5%	198	78.3%	264	55.3%	172	68.0%	197	54.6%	128	65.0%	0.53[3]

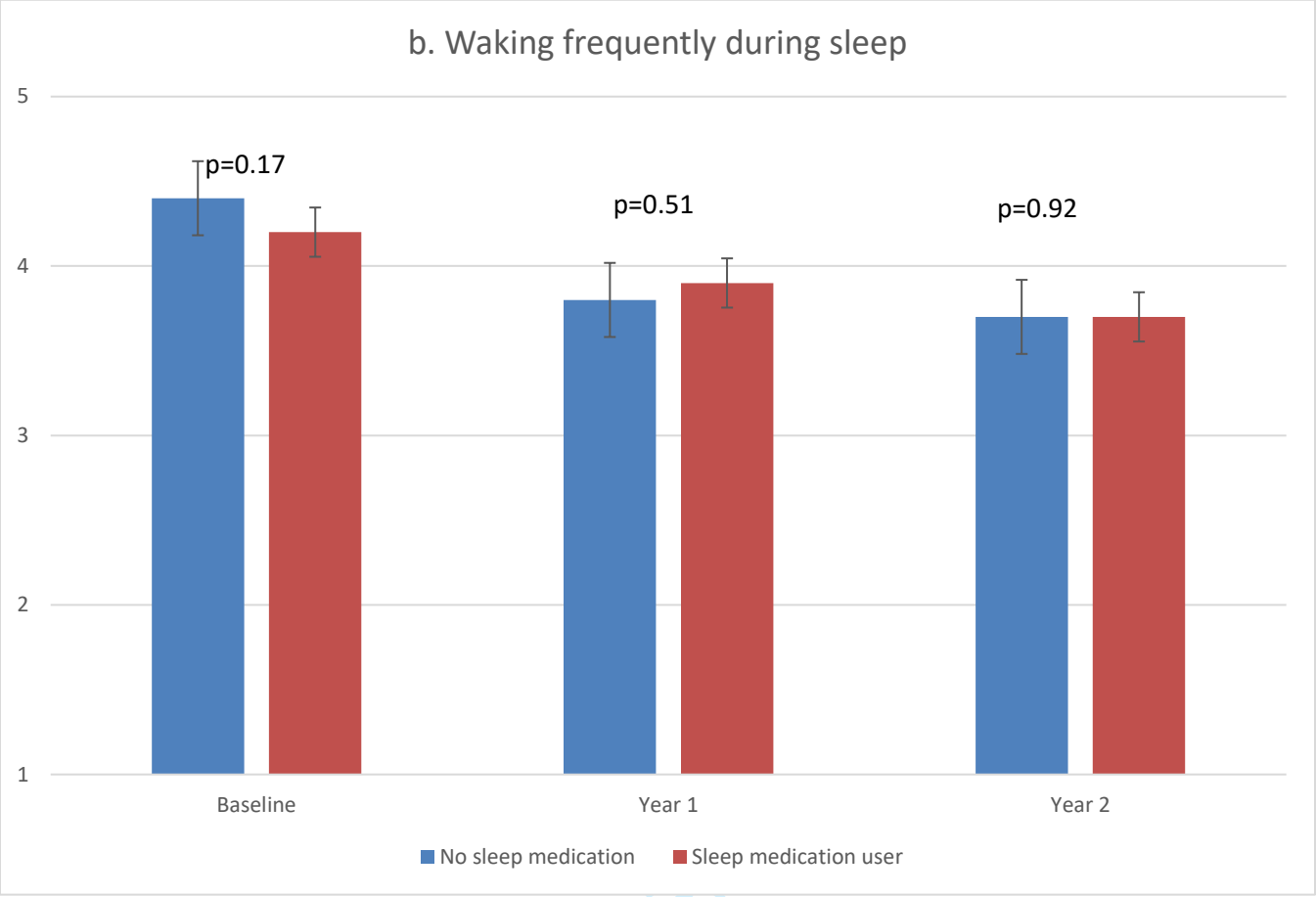
5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on less <1 night/week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week.

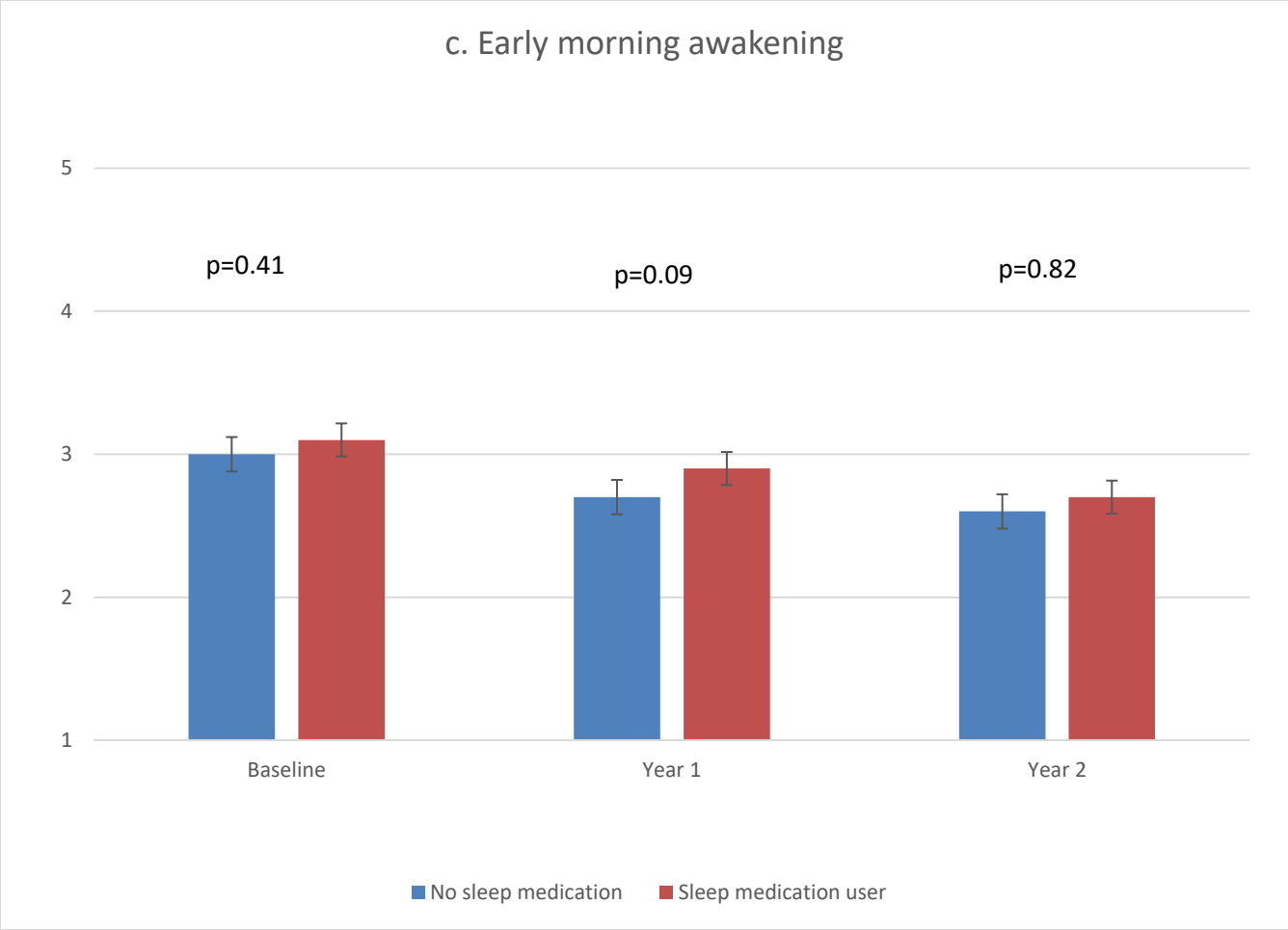
*p-values reflect change in severity of disturbances: [1]=baseline vs visit 1, [2]= baseline vs visit 2, [3]= visit 1 vs. visit 2.

Supplementary Figure 2. Sleep Disturbance Ratings by Medication Exposure During Follow-up In women who reported a 4 or 5 on any severity scale

Legend: Means calculated based on 5-point Likert scale where 1 = no difficulties on any nights, 2 = difficulties on <1 night/week, 3 = 1-2 nights per week, 4 = 3-4 nights per week, and 5 = 5-7 nights per week . Error bars represent standard errors. P-values estimated from the Wilcoxon Rank Sum test.







Supplementary Table 5: Change in severity of sleep disturbances from baseline to year 1 using a proportional odds analysis

	Year 0 versus Year 1	Year 1 versus Year 2	Year 0 versus Year 2
	Odds ratio (95% CI) comparing sleep medication users with non-users		
Difficulty initiating sleep	0.89 (0.58 – 1.30)	0.72 (0.42 – 1.30)	0.80 (0.61 – 1.00)
Waking frequently	0.92 (0.58 – 1.40)	1.44 (0.82 – 2.50)	1.20 (0.90 – 1.50)
Early morning awakening	0.75 (0.49 – 1.20)	1.20 (0.67 – 2.00)	0.94 (0.72 – 1.20)
Notes: The odds ratios represent the odds of a one level increase in the Likert scale, comparing sleep medication users to non-users.			

Supplementary Table 6: Original models additionally adjusted for site and estrogen use

	Year 0 vs Year 1		**Year 0 vs Year 2**		**Year 1 vs Year 2**	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Difficulty initiating sleep	-0.15	0.3	-0.18	0.04	-0.24	0.21
Waking Frequently	-0.03	0.83	0.11	0.22	0.21	0.02
Early morning awakening	-0.19	0.22	-0.12	0.17	-0.07	0.45
The significant p-value in Year 0 vs Year 2 shows that the non users decreased by .12 and the med users increased by .06. The significant p-value in Year 1 vs Year 2 shows that non users decreased by .004 and the med users decreased by .22.						