Clinical Practice Guideline for the Pharmacologic Treatment of Chronic Insomnia in Adults

An American Academy of Sleep Medicine Clinical Practice Guideline

Introduction: The purpose of this guideline is to establish clinical practice recommendations for the pharmacologic treatment of chronic insomnia in adults.

Methods: The American Academy of Sleep Medicine commissioned a task force of five experts in sleep medicine. A systematic review was conducted to identify randomized controlled trials, and the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) process was used to assess the evidence. The task force developed recommendations and assigned strengths based on the quality of evidence, the balance of benefits and harms, and patient values and preferences. Literature reviews are provided for pharmacologic agents for which insufficient evidence was available to establish recommendations. The AASM Board of Directors approved the final recommendations.

Recommendations:
1. We suggest that clinicians use ramelteon as a treatment for sleep-onset insomnia (versus no treatment) in adults. (WEAK FOR)
2. We suggest that clinicians use eszopiclone as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. (WEAK FOR)
3. We suggest that clinicians use zaleplon as a treatment for sleep initiation insomnia (versus no treatment) in adults. (WEAK FOR)
4. We suggest that clinicians use zolpidem as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. (WEAK FOR)
5. We suggest that clinicians use triazolam as a treatment for sleep onset insomnia (versus no treatment) in adults. (WEAK FOR)
6. We suggest that clinicians use temazepam as a treatment for sleep initiation and sleep maintenance insomnia (versus no treatment) in adults. (WEAK FOR)
7. We suggest that clinicians use doxepin as a treatment for sleep maintenance insomnia (versus no treatment) in adults. (WEAK FOR)
8. We suggest that clinicians not use trazodone as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)
9. We suggest that clinicians not use tiagabine as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)
10. We suggest that clinicians not use diphenhydramine as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)
11. We suggest that clinicians not use melatonin as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)
12. We suggest that clinicians not use tryptophan as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)
13. We suggest that clinicians not use valerian as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. (WEAK AGAINST)

A STRONG recommendation is one that clinicians should, under most circumstances, follow. A WEAK recommendation reflects a lower degree of certainty in the appropriateness of the patient-care strategy for all patients. The ultimate judgment regarding propriety of any specific care must be made by the clinician, in light of the individual circumstances presented by the patient, available diagnostic tools, accessible treatment options, and resources.
### Table of Contents

1.0 Introduction
2.0 Background
3.0 Methodology
4.0 Clinical Practice Recommendations
  4.1 Melatonin agonists
  4.2 BZD receptor agonists
  4.3 Benzodiazepines
  4.4 Heterocyclics
  4.5 Anticonvulsants
  4.6 Over-the-counter preparations

5.0 Literature Reviews

6.0 Discussion & Future Directions

---

### 1.0 INTRODUCTION

#### 1.1 Clinical Guidelines and Practice Parameters

The American Academy of Sleep Medicine (AASM) has issued several guidelines, reviews, and practice parameters related to the assessment and management of insomnia. A 2000 review and practice parameter paper addressed the comprehensive evaluation of chronic insomnia. Non-pharmacological management of insomnia has been the subject of two practice parameter papers. No formal, evidence-based standards of practice for pharmacological treatment of insomnia have been published, although clinical guidelines addressing this topic have been issued by various groups. The Standards of Practice Committee of the AASM addressed non-prescription treatments for insomnia in a 2006 paper which concluded that there is sparse or little evidence to support use of these agents for insomnia. Preliminary but conflicting evidence was noted for valerian and first-generation H₁ antagonists for short-term use. A 2005 National Institutes of Health consensus conference on manifestations and management of chronic insomnia found moderate to high-grade evidence to support the efficacy of both cognitive-behavioral therapies and benzodiazepine agonists in the short-term management of insomnia, although noted a relative paucity of data concerning long-term usage of such treatments, despite the chronicity of the condition. Little evidence supporting efficacy of other widely used treatments (sedating antidepressants and non-prescription agents) was found.

A 2008 AASM clinical guideline paper on the evaluation and management of chronic insomnia defined psychological and behavioral therapies as a standard of treatment (the highest level of recommendation at that time). No specific level of recommendation is offered for pharmacological therapies, but the consensus recommendation is that such treatment, when used, should be accompanied by cognitive-behavioral therapies whenever possible. Short/intermediate acting benzodiazepine receptor agonists (benzodiazepines (BZDs) or newer BZD receptor agonistic modulators [BzRAs]) or ramelteon are recommended as first-line pharmacotherapy. Other drugs, such as sedating antidepressants or anticonvulsant medications are recommended as second or third-line agents, particularly when comorbidities (e.g. mood disorder or epilepsy) are present. Other, non-prescription drugs such as over-the-
counter, antihistaminergic sleeping aids and herbal/nutritional agents are not recommended due to lack of demonstrated efficacy as well as safety concerns.

A consensus statement from the British Association for Psychopharmacology\textsuperscript{10} assessed evidence related to chronic insomnia, including management issues, and came to similar conclusions. CBT interventions were recommended as first-line treatment. BzRAs were found effective for short-term use although degradation of improvement following discontinuation of hypnotic was noted to be of concern. Limited evidence and toxicity concerns were cited for other prescription and non-prescription agents although prolonged-release melatonin was recommended as a first-line treatment for insomnia in persons over 55 years-old.

1.2 Aims
This investigation was initiated at the request of the board of directors of the AASM, who also reviewed this document and provided feedback. As noted, no formal clinical practice guidelines for the pharmacological treatment of insomnia have yet been issued by the AASM, despite the fact that this remains, by far, the most common approach to therapy, after treatment of comorbidities. As noted, pharmacotherapy is but one of two possible approaches to treatment, the alternative being cognitive-behavioral therapies for insomnia (CBT-I), already identified as a standard of treatment. While this paper does not directly address the relative benefits of these two approaches, the conclusions and recommendations regarding pharmacotherapy must be considered within the context of specific treatment goals, comorbidities, prior treatment responses, availability, safety, patient preference and cost considerations. Despite the clearly favorable benefit to risk ratio of CBT-I, not all patients with an insomnia disorder can and will derive benefit from this treatment alone. This failure may result from inability to access such treatment (due to availability, cost restraints, etc.), inability or unwillingness to participate in the therapy (e.g. developmentally disabled or infirm), or treatment nonresponsiveness. Thus, pharmacotherapy, alone or in combination with CBT-I, must continue to be considered a part of the therapeutic armamentarium, as it currently is for perhaps 25% of the population. Unfortunately, the information presented below underscores the fact that the vast majority of this group are utilizing medications or substances which are not demonstrated to be effective in managing insomnia and/or have significant potential for harm. For the estimated 3.5-7% of individuals receiving prescription medication for sleep disturbance,\textsuperscript{11-13} significant knowledge gaps and anxieties about the proper usage of these agents exists among the prescribers.

This paper includes a systematic review and meta-analyses which provides the basis of the initial AASM clinical practice guideline for pharmacological management of insomnia. The aims of the present analysis are: 1) to determine the efficacy of prescription and non-prescription medications for treatment of insomnia; 2) to assess the efficacy of individual medications for specific sleep complaints (i.e. difficulty initiating sleep/difficulty maintaining sleep); 3) to evaluate the potential for adverse effects of these drugs; 4) to consider the evidence concerning efficacy and adverse effects in delineating evidence-based guidelines for the use of pharmacotherapy in the management of chronic insomnia.

2.0 BACKGROUND
Insomnia disorder is defined in the International Classification of Sleep Disorders (ICSD-3)\textsuperscript{14} as a complaint of trouble initiating or maintaining sleep which is associated with daytime consequences and not attributable to environmental circumstances or inadequate opportunity to sleep. Beyond the necessary treatment of comorbidities which may be associated with the insomnia disorder, the major treatment interventions fall into two primary categories. Non-pharmacological therapies, largely cognitive behavioral in nature, have been the subject of numerous meta-analyses and practice guidelines.\textsuperscript{5, 15-18} Pharmacological
therapy, including over-the-counter sleep aids and alcohol, is the most widely used treatment for insomnia, yet no evidence-based clinical practice guidelines have been published to date by the American Academy of Sleep Medicine (AASM). This paper includes a systematic review and meta-analyses which provides the basis of the initial AASM clinical practice guideline for pharmacological management of insomnia.

Occasional, short-term insomnia affects 30-50% of the population. The prevalence of chronic insomnia disorder in industrialized nations is estimated to be at least 5-10%. In medically and psychiatrically ill populations, as well as in older age groups, the prevalence is significantly higher. Chronic insomnia is associated with numerous adverse effects on function, health, and quality of life. Epidemiologic studies demonstrate marked impairment in functional status among those with chronic insomnia. Increased rates of work absenteeism, and occupational and motor vehicle accidents have also been widely reported. Persistent insomnia has been identified in multiple studies as a significant risk factor for the development of psychiatric disorders, especially mood disorder. This condition is also associated with increased risk of relapse for depression and alcoholism, as well as adverse effects in chronic pain populations. More recent investigations suggest that chronic insomnia is associated with increased risk for cardiovascular disease. In particular, insomnia with objectively short sleep time is a significant risk factor for the development of hypertension.

Chronic insomnia imposes substantial economic burdens on society. Estimation of direct and indirect costs of chronic insomnia are complicated by many confounding variables, but virtually all analyses of these costs indicate substantially higher economic burden for an insomnia population. Direct cost analysis demonstrates significantly higher utilization of emergency and office health care visits as well as greater cost for prescription drugs. Likewise, indirect costs in the form of work absenteeism, loss of productivity, and insomnia-related accidents contribute significantly to the economic burden of the disorder. In the United States, a 2009 study found that direct and indirect costs for insomnia patients were in excess of $2000/year greater than those of a matched non-insomnia group. Total direct and indirect cost estimates for insomnia in the United States differ substantially due to variability in methodologies. Nevertheless, estimates suggest direct costs of $2-16 billion per year and indirect costs of $75-100 billion. The latter are accounted for in large part by worker absenteeism, presenteeism (lower productivity due to daytime impairment) and work-related accidents.

2.1 History of hypnotic usage
Pharmacological agents have been used for the treatment of insomnia throughout much of recorded history. Prior to the 20th century, opioids, various herbal preparations, bromide salts and alcohol were the primary hypnotic alternatives. Through the first half of the 20th century, barbiturate and related compounds became the most commonly used agents for management of anxiety and sleep disturbance, as well as epilepsy. By mid-century, however, the adverse side effects and lethal overdose potential of these agents became recognized, contributing to curtailment of use.

The first benzodiazepine (BZD) (chlordiazepoxide) was introduced to the U.S. market in 1963, followed shortly by diazepam. Flurazepam, the first benzodiazepine approved by the Food and Drug Administration (FDA) as a hypnotic, became available in 1970 and rapidly supplanted the use of barbiturates and similar compounds for treatment of insomnia. Zolpidem, the first U.S. nonbenzodiazepine, benzodiazepine receptor agonist (non-BZD or BzRA) hypnotic, became available in 1992 and remains the most widely prescribed hypnotic medication, accounting for 87.5% of all BzRA prescriptions in a recent survey of hypnotic use. Since 2005, a melatonin agonist (ramelteon), a low dose sedating tricyclic medication (doxepin), and, most recently, an orexin antagonist (suvorexant) have entered the U.S. market.
2.2 Current Hypnotic Usage

Hypnotic prescribing practices have varied in recent decades as availability of various agents and safety concerns have evolved. Despite the development of numerous BZD hypnotic medications of varying durations of action, the overall frequency of hypnotic prescriptions for insomnia declined during the two decades from 1970-1990, from 3.5 to 2.5%. Due to apparent concerns regarding the potential for tolerance and dependency with BZD use, physicians increasingly prescribed sedating antidepressants “off label,” especially trazodone, despite the absence of efficacy studies for this or any other sedating antidepressants for treatment of insomnia. Survey of office-based physician prescribing practices for the period 1987-1996 revealed an over 50% decline in BZD hypnotic prescriptions accompanied by a nearly 150% increase in trazodone prescriptions. Overall prescriptions for insomnia declined by about 25% during this period. A more recent study, utilizing the National Health and Nutrition Examination Survey (NHANES) data from 1999-2010, analyzed the frequency of usage of medications commonly prescribed for insomnia (MCUFI). This includes BZDs approved for treatment of insomnia, BzRAs, ramelteon, trazodone, doxepin and quetiapine. The authors report that just under 3% of the sample population used one of these agents within the past month. In contrast to the apparent trends of preceding decades, frequency of usage of any MCUFI increased over the decade from 2.0% in the first year sampled to 3.5% in the final year (2009-10). BzRAs, predominantly zolpidem, were most commonly prescribed (1.23% of the population), followed by trazodone (.97%), BZDs (0.4%), quetiapine (0.32%) and doxepin. However, it should be noted in this and other studies that other agents, especially BZDs not approved for insomnia, other antidepressants, antipsychotics, and analgesics, are not included in these data. It seems likely that the true prevalence of medication use for sleep disturbance is higher that these figures suggest. This may explain the fact that the 2005 National Sleep Foundation’s (NSF) survey of sleep habits found that 7% of respondents report using a prescription medication to improve sleep at least a few nights per month.

Physicians and other health care providers have consistently expressed reservations about the use of medication, particularly BZDs and BzRAs, to treat insomnia. They cite concerns regarding safety and dependency as key issues. However, they also note a lack of awareness and/or availability of alternative treatments. Many favor an initial approach of treating associated comorbidities and advising good sleep hygiene. An ever-increasing amount of data makes it clear that the latter approach is very often unsuccessful, leaving providers feeling compelled to prescribe medications. Most of those surveyed recognize the need for additional, non-pharmacological treatment for their patients, but cite a number of barriers to acquiring such treatment.

Data concerning use of non-prescription agents for sleep promotion is limited. The aforementioned NSF survey reported that nearly one in four respondents use some form of sleep aid “at least several times per month.” Eleven percent stated that they used alcohol to help sleep at this frequency; 9% used over-the-counter sleep aids and 2% used melatonin.

2.3 Previous Meta-analyses

Several meta-analyses of pharmacotherapy for insomnia have been conducted. Nowell and colleagues conducted a meta-analysis of 22 randomized controlled trials (RCTs) of BZD's and zolpidem for treatment of primary insomnia published from 1966 to 1996. They found moderate effect sizes (d=0.56 to 0.71 for key sleep variables) for improvement with these agents, but noted limitations in the duration of trials and lack of follow-up study regarding outcome. A 2000 study commissioned by the Canadian Medical Association evaluated 45 RCT's (N=2672) of BZDs for treatment of primary insomnia. This investigation found reduction in sleep latency (non-significant in objective [PSG] assessment but significant in subjective reports) and a somewhat larger and significant increase in total sleep time by both objective and subjective reports. The authors also note an increase in adverse events with BZDs (pooled odds ratio for any adverse event = 1.8) and call into question the risk/benefit ratio for these agents.
A comparative evaluation of the efficacy of hypnotic drugs was conducted by the National Centre for Clinical Excellence of the UK. In summary, the analysis found little difference among the numerous BZDs and BzRAs included in the 24 studies which compared a BZD to a BzRA or two BzRAs to one another. Some small differences were noted such as shorter sleep latency but less total sleep time with zaleplon when compared to zolpidem. On the whole, major differences in adverse effects were not observed between drugs. Meta-analyses in this report were few due to limitations of data reporting.

Glass and colleagues compared benefits versus risks for all sedative hypnotic agents in a meta-analysis of RCTs of active agent vs. placebo or other active compound in populations >60 years of age free of contributing comorbidities. They reported a small effect size for improvement in sleep quality (d=0.14). Separate analysis of BZDs alone yielded a somewhat more robust improvement in quality (d=.37). Significant but modest increase in total sleep time and reduction in number of awakenings were also found for all sedative-hypnotics and for the BZD group alone, although effect sizes are not reported for these variables. Cognitive side effects were more common with sedative-hypnotics. The authors note that with respect to the sleep quality measures reported for all sedative hypnotics, the number needed to treat is 13, while the number needed to harm is 6, thereby indicating an unfavorable risk/benefit ratio for this population. Independent analysis of this ratio for BZDs alone was not conducted.

A 2007 meta-analysis evaluated 105 RCTs of BZDs, BzRAs and antidepressant medications for treatment of chronic insomnia in the adult populations regardless of comorbidities. In summary, the analysis indicates moderate and significant improvement in major sleep parameters with both BZDs and BzRAs in both objective (PSG) and subjective (sleep diary) assessments with the exception of PSG results for WASO and TST, which yielded results just below the range of significance. Far fewer studies were available for antidepressants. These showed significant reduction in sleep latency and a non-significant trend toward reduced WASO. Four studies utilizing PSG measures showed substantial improvement in TST (79.6m) while single subjective data set suggested reduction in TST compared to placebo. The authors note substantial heterogeneity of data which was reduced in subgroup analyses by type of drug. Between groups comparisons showed no significant efficacy differences between BZDs and non-BZDs. All three groups demonstrated significantly higher rates of adverse events vs. placebo. BZDs exhibited a higher rate of adverse events than BzRAs.

Huedo-Medina and colleagues conducted systematic review and meta-analysis of data on BzRAs submitted to the U.S. Food and Drug Administration from 15 studies. They found that BzRAs produce significant reduction of sleep latency by both objective and subjective measures with effect sizes of 0.36 and 0.33, respectively. Other sleep variables did not show significant differences but limited data reporting on these variables precluded definitive conclusions.

Finally, Winkler and Doering analyzed data from 31 randomized controlled trials of sleep-promoting substances used for treatment of primary insomnia. Studies included BZDs, BzRAs, melatonin agonists, antidepressants and other sedating compounds. Only studies which included objective (polysomnographic) data were considered. The meta-analysis revealed that both BZDs and BzRAs were significantly more effective than antidepressants. Both demonstrated small to moderate effect sizes for major sleep variables. BZDs were somewhat superior to BzRAs for subjective SL. No analysis of treatment-emergent adverse events was reported.
3.0 METHODS

3.1 Expert Task Force
In order to develop this clinical practice guideline, the AASM commissioned a task force (TF) composed of five content experts in the field of insomnia, an AASM BOD liaison, and AASM Science and Research Department staff members. Prior to appointment, the content experts were required to disclose all potential conflicts of interest (COI) according to the AASM’s policy. All relevant conflicts of interest are listed in the Disclosures section.

3.2 PICO Questions
A PICO (Patient, Population or Problem, Intervention, Comparison, and Outcomes) question template was developed to be the focus of this guideline:

“In adult patients diagnosed with primary chronic insomnia, how does [intervention] improve [outcomes], compared to placebo?”

The PICO question template was approved by the AASM Board of Directors. The TF identified the pharmacological interventions of interest, based on FDA approval status and common off-label usage. The TF then developed a list of patient-oriented clinically relevant outcomes that are indicative of whether a treatment should be recommended for clinical practice. The TF rated their relative importance and selected the top six outcomes. Sleep Latency, Wake After Sleep Onset, Total Sleep Time, Quality of Sleep, Number of Awakenings, and Sleep Efficiency were determined to be “critical” or “important” for clinical decision making across all interventions (Table 1). Clinical Significance Thresholds, used to determine if a change in an outcome was clinically significant, were defined for each outcome by TF clinical judgement, prior to statistical analysis (Table 2). Lastly, The TF determined which outcomes were “critical” for clinical decision making for each individual intervention (Table 3).

<table>
<thead>
<tr>
<th>Table 1- PICO Question Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPULATION</td>
</tr>
<tr>
<td>Adult patients diagnosed with</td>
</tr>
<tr>
<td>primary chronic insomnia</td>
</tr>
<tr>
<td>INTERVENTION</td>
</tr>
<tr>
<td>1. Diphenhydramine</td>
</tr>
<tr>
<td>2. Doxepin</td>
</tr>
<tr>
<td>3. Eszopiclone</td>
</tr>
<tr>
<td>4. Melatonin</td>
</tr>
<tr>
<td>5. Ramelteon</td>
</tr>
<tr>
<td>6. Temazepam</td>
</tr>
<tr>
<td>7. Tiagabine</td>
</tr>
<tr>
<td>8. Trazadone</td>
</tr>
<tr>
<td>9. Triazolam</td>
</tr>
<tr>
<td>10. Tryptophan</td>
</tr>
<tr>
<td>11. Valerian-hops</td>
</tr>
<tr>
<td>12. Zaleplon</td>
</tr>
<tr>
<td>13. Zolpidem</td>
</tr>
<tr>
<td>COMPARISON</td>
</tr>
<tr>
<td>Placebo control</td>
</tr>
<tr>
<td>OUTCOMES</td>
</tr>
<tr>
<td>Sleep Latency (SL)</td>
</tr>
<tr>
<td>Total Sleep time (TST)</td>
</tr>
<tr>
<td>Wake After Sleep Onset (WASO)</td>
</tr>
<tr>
<td>Quality of Sleep (QOS)</td>
</tr>
<tr>
<td>Sleep Efficiency (SE)</td>
</tr>
<tr>
<td>Number of Awakenings (NOA)</td>
</tr>
</tbody>
</table>
Table 2 - “Critical” Outcomes by Intervention

<table>
<thead>
<tr>
<th>Intervention</th>
<th>TST</th>
<th>SL</th>
<th>WASO</th>
<th>QOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diphenhydramine</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Doxepin</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Eszopiclone</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Melatonin</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ramelteon</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Temazepam</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tiagabine</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trazadone</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Triazolam</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Valerian-hops</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zaleplon</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zolpidem</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 3 - Clinical Significance Threshold

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Measurement Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep Latency (SL) (min)</td>
<td>PSG</td>
</tr>
<tr>
<td>Total Sleep Time (TST) (min)</td>
<td>Actigraphy</td>
</tr>
<tr>
<td>Wake after Sleep Onset (WASO) (min)</td>
<td>Subjective</td>
</tr>
<tr>
<td>Quality of Sleep (QOS) (varies)</td>
<td>varies</td>
</tr>
<tr>
<td>Sleep Efficiency (SE) (%)</td>
<td>varies</td>
</tr>
<tr>
<td>Number of Awakenings (NOA) (number)</td>
<td>varies</td>
</tr>
</tbody>
</table>

3.3 Literature Searches, Evidence Review and Data Extraction

Multiple literature searches were performed by the AASM research staff using the PubMed database throughout the guideline development process (see Figure 1). Keywords and MeSH terms were:

- Insomnia OR Sleep Initiation and Maintenance Disorder NOT Transient AND
- Clinical trial OR randomized controlled trial
- NOT Editorial, Letter, Comment, Case Reports, Biography, Review

The full literature search string can be found in the supplemental materials (see Supplemental Materials, “Literature Search Terms”). Searches were performed on April 26, 2011 (search 1), May 12, 2014 (search 2), and October 15, 2014 (search 3). The “pearling” process served as a “spot check” for the literature searches, to ensure that important articles were not missed. Based on their expertise and familiarity with the insomnia literature, TF members submitted additional relevant literature, and screened reference lists to identify articles of potential interest.

Abstracts from all retrieved articles, including “pearled” publications, were individually assessed by two TF members to determine whether the publication should be included or excluded from further consideration in the project. Exclusion criteria can be found in Figure 1 and the supplemental materials (see
Supplemental Materials, "Inclusion/Exclusion criteria"). This resulted in 127 publications being approved for inclusion.

Full texts of accepted articles were inspected closely; data pertaining to GRADE for the outcomes of interest were extracted into spreadsheets by AASM staff. All data pertaining to adverse events were extracted into separate spreadsheets. Articles that met inclusion criteria, but did not report outcomes of interest were rejected from the final evidence base. If outcome data were not presented in the format necessary for statistical analysis (i.e. mean, standard deviation, and sample size), the authors were contacted in an attempt to obtain the necessary data. If the necessary data were not available, the paper was included in the evidence base as supporting evidence, but was not used for statistical analysis or for determining the quality of evidence. Of the 127 accepted publications, 44 were included in the statistical and meta-analysis.

Figure 1 - Evidence Base Flow Diagram

2813 studies identified through PubMed.
Search 1: all publications prior to April 26, 2011.
Search 2: April 1, 2011 to May 12, 2014
Search 3: April 1, 2014 to October 15, 2014
Search 4: October 1, 2014 to January 25, 2016

2821 studies screened for inclusion/exclusion criteria

2697 studies excluded.
Reason for exclusion:
  a. Not a drug treatment
  b. Pediatric
  c. Initial sample size < 20
  d. Wrong publication type (review, editorial, methodological, etc.)
  e. Significant comorbidity
  f. Major topic not chronic insomnia
  g. Normal/healthy subjects
  h. Other subpopulation (hospitalized pts, intellectual disabilities, etc.)

8 studies identified through pearling

127 studies included in evidence base for recommendations

44 studies included in meta-analysis

If none of the accepted publications for an intervention provided data that could be used for statistical analysis, the task force did not make a recommendation and provided a literature review of these accepted papers instead. These publications are not included in Figure 1.

3.4 Statistical and Meta-Analysis
For outcomes of interest, data from baseline and last-treatment time points were used for all statistical and meta-analyses. Analyses were limited to only FDA-approved doses. For adverse events, all data presented
in the papers included in the evidence base were used for statistical and meta-analysis. All calculations and meta-analyses were performed using Review Manager 5.3 software. Whenever possible, meta-analyses were performed by pooling data across studies for each outcome and adverse event. The evidence was grouped for analysis based on the drug, dosage, clinical outcome of interest, and methodology used to obtain the data (e.g., data obtained by PSG were analyzed separately from data obtained by sleep diary).

All meta-analyses were performed as per-treatment analyses using the random effects model. For most interventions, absolute effects of drug treatments are represented by the mean difference (MD) ± standard deviation (SD) of post-treatment vs post-placebo. Meta-analyses for adverse events are presented as risk difference. The result of each meta-analysis is displayed as a forest plot. Pooled results are expressed as the total number of patients, mean difference (MD) and 95% confidence interval (CI) between the experimental treatment and placebo.

Interpretation of clinical significance for outcomes of interest was conducted by comparing the absolute effects of drug treatment to the Clinical Significance Threshold previously determined by the TF for each outcome of interest. Interpretation of adverse events was based upon the risk difference and clinical expertise of the task force.

### 3.5 Strength of Recommendations

The GRADE approach (Grades of Recommendation, Assessment, Development and Evaluation) was used for the assessment of quality of evidence. The TF assessed the following three components to determine the direction and strength of a recommendation: quality of evidence, balance of beneficial and harmful effects, and patient values and preferences.

For the determination of the quality of evidence for an intervention, the TF used objective data whenever possible (e.g., polysomnography). When only subjective data were available (e.g., sleep diaries), this evidence was used to determine the overall Quality of Evidence. The results of this assessment are presented as Summary of Findings (SoF) tables for each intervention (see Supplemental Materials, Table s1-s21).

The task force developed recommendation statements and determined strengths of these recommendations based on the balance of the following factors:

1. Quality of evidence – based on an assessment of the overall risk of bias (randomization, blinding, allocation concealment, selective reporting, and author disclosures), imprecision (clinical significance thresholds), inconsistency (I² cutoff of 75%), indirectness (study population), and risk of publication bias (funding sources), the task force determined their overall confidence that the estimated effect found in the body of evidence was representative of the true treatment effect that patients would see.

2. Benefits vs. Harms – based on the meta-analysis (if applicable), analysis of any harms/side effects reported within the accepted literature, and the clinical expertise of the task force, the task force determined if the beneficial outcomes of the intervention outweighed any harmful side effects.

3. Patient values and preferences – based on the clinical expertise of the task force members and any data published on the topic relevant to patient preferences, the task force determined if patient values and preferences would be generally consistent, and if patients would use the intervention based on the body of evidence.

Taking these major factors into consideration, each recommendation statement was assigned a direction (For or Against) and a strength (Strong or Weak).
Additional information is provided in the form of “Remarks” immediately following the recommendation statements, when deemed necessary by the task force. Remarks are based on the evidence evaluated during the systematic review, and are intended to provide context for the recommendations.

3.6 Approval and Interpretation of Recommendations
A draft of the guideline was made available for public comment for a two-week period on the AASM website. The task force took into consideration all the comments received and made revisions when appropriate. The final guideline was submitted to the AASM Board of Directors who approved these recommendations.

The recommendations in this guideline define principles of practice that should meet the needs of most patients in most situations. This guideline should not, however, be considered inclusive of all proper methods of care or exclusive of other methods of care reasonably used to obtain the same results. A STRONG recommendation is one that clinicians should, under most circumstances, always be doing (i.e. something that might qualify as a Quality Measure). A WEAK recommendation reflects a lower degree of certainty in the appropriateness of the patient-care strategy and requires that the clinician use their clinical knowledge and experience, and refers to the individual patient's values and preferences to determine the best course of action. The ultimate judgment regarding propriety of any specific care must be made by the clinician, in light of the individual circumstances presented by the patient, available diagnostic tools, accessible treatment options, and resources.

The AASM expects this guideline to have an impact on professional behavior, patient outcomes, and, possibly, health care costs. This clinical practice guideline reflects the state of knowledge at the time of publication and will be reviewed and updated as new information becomes available.

4.0 CLINICAL PRACTICE RECOMMENDATIONS
The following clinical practice recommendations are based on the systematic review and evaluation of evidence following the GRADE methodology. Remarks are intended to provide context for the recommendations. A summary of the recommendations and GRADE determinations is presented in Table 4.

| It is essential that the recommendations which follow be interpreted within the appropriate context of clinical practice. Readers will note that all specific recommendations fall within the “weak” (for or against) classification of the GRADE system. This should not be construed to mean that there are no sleep-promoting medications that are clearly efficacious or indicated in the treatment of chronic insomnia. Hypnotic medications, along with management of comorbidities and non-pharmacological interventions such as CBT, are an important therapeutic option for chronic insomnia. The strength of recommendations within the GRADE system are driven by the degree of confidence in a variety of factors related to the intervention including 1) the availability of specific data regarding efficacy; 2) the quality of that data, and 3) other considerations such as potential risks, impact of treatment, patient values and preferences, and perceived burden of treatment. The existing data regarding sleep-promoting medications imposes limits on the degree of confidence as a result of several factors. These include: 1) a high degree of variability in the |

11
statistical information presented. Many studies, especially older studies, do not present results that meet the criteria for meta-analysis within GRADE and are, by necessity, excluded from formal analysis; 2) a significant degree of variability in sleep outcomes within and across studies. Such variability produces a “downgrading” of the quality of evidence within GRADE; 3) industry sponsorship. There is very limited funding for clinical trials outside of industry. As a result, the quality of evidence for a vast majority of available data is downgraded due to potential publication bias associated with such sponsorship; 4) a paucity of systematic data collection and analysis of treatment-emergent adverse events. Absent such information, it is difficult to assign a high degree of confidence to estimations of benefit:risks ratio; and 5) absence of outcome data (such as functional status or prevention of complications of chronic insomnia) that would inform judgments regarding the impact of therapy.

The strength (or weakness) of these recommendations speaks as much, or more, to the limitations of the data as they do to the relative benefits and risks of the treatments per se. Clinicians must continue to exercise appropriate judgment, based not only on the recommendations presented herein, but also on individual patient characteristics, comorbidities and patient preferences in the prescribing of sleep-promoting medications and general management of chronic insomnia.

Finally, the literature review, meta-analyses and recommendations are based only on FDA-approved doses. This should not be interpreted as a recommendation for the use of a specific dose in clinical practice. Numerous factors, including, but not limited to age, comorbidities, and concurrent use of other medications may affect dosage recommendations. Clinical judgment is necessary in determining appropriate dosage, on a patient-by-patient basis.

Table 4 – Summary of clinical practice recommendations

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Recommendation</th>
<th>Direction and Strength of Recommendation</th>
<th>Quality of Evidence</th>
<th>Benefits &amp; Harms assessment</th>
<th>Patients’ Values and Preferences Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Melatonin agonists</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ramelteon</strong></td>
<td>4.1a We suggest that clinicians use ramelteon as a treatment for sleep-onset insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>VERY LOW</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>BZD receptor agonists</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eszopiclone</strong></td>
<td>4.2a We suggest that clinicians use eszopiclone as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>LOW</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Zaleplon</strong></td>
<td>4.3a We suggest that clinicians use zaleplon as a treatment for sleep initiation insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>LOW</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Zolpidem</strong></td>
<td>4.4a We suggest that clinicians use zolpidem as a treatment for sleep-onset and sleep maintenance</td>
<td>WEAK FOR</td>
<td>VERY LOW</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td>Drug</td>
<td>Recommendation Details</td>
<td>Benefits vs Harms</td>
<td>Majority of Patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Triazolam</strong></td>
<td>4.5a We suggest that clinicians use triazolam as a treatment for sleep onset insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>HIGH</td>
<td>Benefits approximately equal to harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Temazepam</strong></td>
<td>4.6a We suggest that clinicians use temazepam as a treatment for sleep initiation and sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>MODERATE</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Doxepin</strong></td>
<td>4.7a We suggest that clinicians use doxepin as a treatment for sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK FOR</td>
<td>LOW</td>
<td>Benefits outweigh harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Trazodone</strong></td>
<td>4.8a We suggest that clinicians not use trazodone as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK AGAINST</td>
<td>MODERATE</td>
<td>Harms outweigh benefits</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Tiagabine</strong></td>
<td>4.9a We suggest that clinicians not use tiagabine as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK AGAINST</td>
<td>VERY LOW</td>
<td>Harms outweigh benefits</td>
<td>The majority of patients would not use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Diphenhydramine</strong></td>
<td>4.10a We suggest that clinicians not use diphenhydramine as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK AGAINST</td>
<td>LOW</td>
<td>Benefits approximately equal to harms</td>
<td>The majority of patients would not use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>Melatonin</strong></td>
<td>4.11a We suggest that clinicians not use melatonin as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults.</td>
<td>WEAK AGAINST</td>
<td>VERY LOW</td>
<td>Benefits approximately equal to harms</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
<tr>
<td><strong>L-tryptophan</strong></td>
<td>4.12a We suggest that clinicians not use tryptophan as a treatment for sleep onset or sleep maintenance insomnia</td>
<td>WEAK AGAINST</td>
<td>MODERATE</td>
<td>Harms outweigh benefits</td>
<td>The majority of patients would use this treatment (over no treatment)</td>
</tr>
</tbody>
</table>
Valerian

This recommendation is based on variable dosages of valerian and valerian-hops combination.

| Valerian | 4.13a We suggest that clinicians not use valerian as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. | WEAK AGAINST | VERY LOW | Benefits approximately equal to harms | The majority of patients would not use this treatment (over no treatment) |

Melatonin agonists

4.1 Ramelteon for the treatment of Primary Insomnia

4.1a We suggest that clinicians use ramelteon as a treatment for sleep-onset insomnia (versus no treatment) in adults. [WEAK FOR]

Remarks: This recommendation is based on Ramelteon 8 mg.

4.1b Summary

Four RCTs investigated the use of ramelteon in the treatment of chronic primary insomnia.\textsuperscript{50-53} The overall quality of evidence from these studies was downgraded to VERY LOW due to substantial heterogeneity across studies, imprecision and potential publication bias. The overall evidence for ramelteon 8 mg was weakly in favor of its effectiveness for the treatment of sleep initiation disturbance only. Meta-analysis of the three studies meeting inclusion criteria which reported objective (PSG) sleep latency demonstrated marginal reduction of sleep latency. The analysis revealed minimal increase in PSG-determined total sleep time which fell well below the defined threshold for clinical significance. Measures of sleep efficiency and sleep quality showed no clinically significant improvement. There was no evidence of significant difference from placebo for any adverse events, based on available side effect data. Although the evidence for efficacy is marginal, the benefits appear to be greater than the minimal potential harms. Based on clinical judgment, the task force determined that the majority of well-informed patients would use ramelteon over no treatment. This judgement is based on the evidence of improved sleep latency, coupled with its apparently low potential for adverse events.

See Supplemental Materials; Figures S1-S8 and S68-S69, and Table S1.

4.1c Discussion

Evidence from four RCTs which investigated the use of ramelteon in the treatment of chronic primary insomnia was included in the statistical analysis.\textsuperscript{50-53} Subjects in all studies demonstrated chronic primary insomnia with associated daytime complaints. All studies required mean objective latencies to persistent sleep (LPS) of >20 minutes on two nights of PSG screening. All studies except Mayer and colleagues also required mean objective wake time after sleep onset (WASO) >60 minutes. Kohsaka and colleagues\textsuperscript{50} studied 65 chronic insomnia patients for two nights each at ramelteon doses of 4, 8, 16, and 32 mg. Roth and colleagues\textsuperscript{52} studied 100 older adults (age >65 years) with chronic primary insomnia. Subjects were administered two consecutive nights each of placebo, ramelteon 4 mg and ramelteon 8 mg in a three phase crossover protocol, with randomization of the treatment sequence and sustained washout time between each two night sequence.

Zammitt\textsuperscript{53} studied the effects of nightly ramelteon in adults at dosages of 8 and 16 mg. PSG was conducted at baseline, and weeks 1, 3, and 5. The Mayer paper\textsuperscript{51} reported on six-month nightly use of ramelteon in
451 adults with chronic insomnia from 46 multinational sites. Two nights of polysomnography were conducted in week 1 and at approximately one month intervals thereafter.

**Sleep latency**
The impact of 8 mg ramelteon on polysomnographically-assessed sleep latency (SL) was evaluated in three studies.\textsuperscript{50, 52, 53} Objective sleep latency data in the study by Mayer and colleagues\textsuperscript{51} were not adequate for meta-analysis and therefore could not be included.

Meta-analysis of the grouped evidence demonstrated marginal improvement in this critical outcome. However, the mean difference between the treatment and control groups was not clinically significant (-9.57 min; CI: -6.38 to -12.75 min). The confidence interval crossed the clinical significance threshold, and therefore the quality of evidence was downgraded for imprecision. It was downgraded further for the high degree of heterogeneity across studies ($I^2=96\%$), and due to the risk of publication bias since all these studies were funded by industry. The resultant quality of evidence is VERY LOW.

Mean differences in objective sleep latency varied from -7.6 minutes to -13.1 minutes. Of note, the Roth investigation\textsuperscript{4} included exclusively older adults and found the smallest improvement in sleep latency. Subjective sleep latency from these investigations was comparable to objective latencies with mean difference (-11.44 min; CI: -3.31 to -19.56 min) falling below the clinical significance threshold.

Several additional papers which met inclusion criteria, but did not contain data suitable for this analysis, have addressed the efficacy and side effect profile of ramelteon.\textsuperscript{54-58} The objective and subjective sleep latency from these results were consistent with the meta-analysis findings. This was likewise the case for sub-group analysis of subjects with primary sleep initiation complaints.\textsuperscript{56} A post-hoc analysis of the data from Zammitt by Mini and colleagues\textsuperscript{55} found a significantly greater percentage of ramelteon 8 mg patients with >50% reduction in sleep latency at week 1 (63.0% vs 39.7% for placebo), week 3 (63.0% vs 41.2%), and week 5 (65.9% vs 48.9%).

**Total sleep time**
All four studies included in the meta-analysis evaluated objective total sleep time for ramelteon 8 mg.\textsuperscript{50-53} Although small improvements in TST were observed in some individual studies, ranging from 1.2 to 12.5 minutes longer, the meta-analysis reveals minimal increase (+6.58 min; CI: +1.36 to +11.80 min) which falls well below the threshold for clinical significance. The quality of evidence was downgraded to LOW due to the high degree of heterogeneity across studies, and due to the risk of publication bias since all these studies were funded by industry. Meta-analysis results of reported subjective TST were consistent with the objective finding (+5.7 min; CI: -7.65 to +19.04 min). Additional studies not included in meta-analysis supported these results.\textsuperscript{54,57,58}

**Wake after sleep onset**
Meta-analysis of objective WASO from the two studies reporting adequate data\textsuperscript{50, 53} show a clinically insignificant increase (+3.5 min; CI: +2.77 to +4.23 min) in WASO for the ramelteon group, well below the significance threshold of 20 minutes. The quality of evidence was downgraded to MODERATE due to potential publication bias. One study not included in meta-analysis (Erman, 2006) found no difference in PSG WASO.

Zammitt and Mayer reported subjective WASO data for meta-analysis.\textsuperscript{51, 53} The ramelteon group demonstrated a clinically insignificant increase in WASO of 5.2 min (CI: -6.77 to +17.24 min). The quality of evidence was LOW due to heterogeneity and potential publication bias. The only additional study which assessed subjective WASO found no difference between placebo and ramelteon 8 mg.\textsuperscript{58}
Quality of sleep
Sleep quality ratings showed virtually no difference from placebo in any of the studies assessed.\textsuperscript{51-53} Meta-analysis suggests no difference between ramelteon and placebo, with a pooled mean difference of -0.04 (CI: -0.13 to +0.05 MD) on a 7-point Likert scale. The quality of evidence was downgraded to LOW due to heterogeneity and the risk of publication bias since all these studies were funded by industry. Additional studies which assessed subjective sleep quality found no difference between ramelteon and placebo groups.\textsuperscript{54, 57, 58}

Sleep efficiency
Three studies reported sleep efficiency data included in meta-analysis.\textsuperscript{50, 52, 53} Minimal improvements in sleep efficiency were reported (+1.93%; CI: +1.00 to +2.87%), falling well below the clinically significance threshold for objective sleep efficiency of 5%. The quality of evidence was LOW due to heterogeneity and potential publication bias. Additional studies did not report sleep efficiency data.

Number of awakenings
No meta-analysis for PSG number of awakenings was conducted as only one study reported adequate data for analysis.\textsuperscript{51} This investigation found no clinically significant difference between ramelteon 8 mg and placebo (mean difference 0.1; CI: +0.08 to +0.15 MD). The quality of evidence was MODERATE due to potential publication bias. Other studies which evaluated NOA reported no significant differences as well.\textsuperscript{52, 57, 58}

In summary, these studies show very weak evidence for reduction of sleep latency at the recommended prescribed dosage (8 mg), with mean decrease of 9.57 minutes (CI: -6.38 to -12.75 min), and no consistent evidence of improvement in other objective or subjective parameters.

Overall Quality of Evidence
The overall quality of evidence in the meta-analytic data from these studies was downgraded to very low for several reasons. Substantial heterogeneity across studies was noted for multiple outcomes. The data were also downgraded for imprecision, due to the relatively large confidence intervals, which cross the clinical significance thresholds for multiple outcomes. All of these studies were industry sponsored, resulting in further downgrading of evidence due to potential publication bias. The quality of evidence for individual outcomes ranged from moderate to very low, therefore the overall quality of evidence was VERY LOW.

Harms
Meta-analytic data on adverse effects showed a relatively low frequency of adverse effects overall and none which were significantly different than placebo. This analysis included headache, nausea, upper respiratory infection and nasopharyngitis. A single case of leukopenia, which was judged possibly related to medication, was noted in the Mayer study.\textsuperscript{51} Both Zammitt and Mayer\textsuperscript{51, 53} found no evidence of rebound insomnia or withdrawal effects following discontinuation (notably, the Mayer et al study was six-months of nightly use).

The studies not included in the meta-analysis found no indication of a significant difference in adverse events between ramelteon and placebo. Commonly reported adverse events in these studies included fatigue, headache, dizziness and somnolence.

Three studies assessed for next-day impairment associated with ramelteon. Roth and colleagues reported on next-day residual pharmacological effects of ramelteon in an older adult population.\textsuperscript{52} Observations of digit symbol substitution (DSST), immediate and delayed recall, subjective alertness and concentration showed no significant residual as compared to placebo on any outcomes. Employing the same residual
effect measures, Zammitt et al\textsuperscript{53} reported small but statistically significant impairment with ramelteon 8 mg. in immediate recall at week 3 only, delayed recall (week 1 only), level of alertness (week 5) and ability to concentrate (week 1). Mayer\textsuperscript{51} found no consistent evidence of next-day impairment in alertness, recall, DSST or visual analogue scales of mood, energy or cognition. Overall, the available data suggest no consistent evidence of next-day impairment associated with the use of ramelteon.

In summary, the task force found that there was weak evidence of efficacy in the treatment of sleep initiation insomnia, with limited or no consistent evidence of adverse events in excess of placebo. Therefore, benefits were deemed to marginally outweigh harms.

**Patients’ Values and Preferences**

Based on its clinical judgement, the task force determined that in light of its efficacy for sleep initiation and its relatively benign side effect profile, a majority of patients would be likely to use ramelteon compared to no treatment.

**BZD receptor agonists**

4.2 Eszopiclone for the Treatment of Primary Insomnia

4.2a We suggest that clinicians use eszopiclone as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. [WEAK FOR]

Remarks: This recommendation is based on eszopiclone 2mg and 3 mg.

4.2b Summary

Six RCTs evaluated eszopiclone 2 mg for the treatment of chronic primary insomnia.\textsuperscript{59-64} The overall quality of evidence was downgraded to LOW due to imprecision and risk of publication bias. The evidence for eszopiclone 2 mg was weakly in favor of its effectiveness for improving sleep initiation disturbance and total sleep time. Meta-analysis data from three studies which reported objective sleep latency showed a clinically significant mean reduction in PSG sleep latency.\textsuperscript{61, 63, 64} Four studies which evaluated subjective total sleep time demonstrated a significant mean increase versus placebo.\textsuperscript{60, 62, 64} Assessment of PSG sleep efficiency in two studies\textsuperscript{61, 64} and subjective sleep quality in four studies\textsuperscript{60, 62-64} revealed improvement which fell just below the threshold for clinical significance. Measures of reduction in wake time after sleep onset and number of awakenings revealed trends toward improvement which fell below the defined level of clinical significance. Meta-analysis of adverse effects, derived from all six studies, indicated no significant differences from placebo.

Six studies assessed the effects of eszopiclone 3 mg for treatment of chronic primary insomnia.\textsuperscript{60, 63-67} The quality of evidence for these studies as a whole was downgraded to VERY LOW due to significant heterogeneity, imprecision and potential publication bias. The collective evidence for eszopiclone 3 mg was weakly in favor of effectiveness for improving sleep initiation, total sleep time, sleep efficiency, number of awakenings and sleep quality. The meta-analysis data from three studies demonstrated clinically significant reduction in objective sleep latency.\textsuperscript{63-65} Four studies likewise revealed clinically significant increase in mean subjective total sleep time.\textsuperscript{60, 64, 66, 67} PSG sleep efficiency, reported in two studies\textsuperscript{64, 65}, also exceeded the threshold for clinically significant improvement, as did subjective sleep quality, which was reported in all six studies included in meta-analysis. A trend in the direction of reduced WASO was observed but did not reach clinical significance. Insufficient data were available for meta-analysis of eszopiclone 3 mg adverse effects.
Overall, the benefits of eszopiclone 2 mg and 3 mg were judged to be greater than the minimal potential harms. Based on clinical judgment, the task force determined that the majority of well-informed patients would use eszopiclone 2 mg and 3 mg over no treatment. This judgement is based on the evidence of improvement in sleep latency, total sleep time, sleep efficiency and sleep quality, coupled with its low potential for adverse events.

See Supplemental Materials; Figures S9-S25 and S70-S71, and Tables S2-S3.

4.2c Discussion
A total of nine studies were included in the meta-analyses for eszopiclone 2 mg and 3 mg.\textsuperscript{59-67} Three of these studies included only older adults (>65 years).\textsuperscript{59, 61, 62} The remainder studied adults, typically 21-65 years of age. Inclusion criteria for most of these studies required persistent subjective sleep latency >30 minutes and TST <6.5 hours.\textsuperscript{60-65} Ancoli-Israel and colleagues\textsuperscript{59} studied 388 older adults for 12 consecutive weeks of nightly eszopiclone 2 mg use. Inclusion criteria for this study specified TST <6 hours and WASO >45 minutes. Outcome data were patient-reported. McCall and colleagues\textsuperscript{61} also reported on two-week administration of 2 mg eszopiclone vs. placebo to 254 older adults. In addition to sleep latency and TST, inclusion criteria, subjects were required to have WASO >20 minutes. PSG was conducted on nights 1, 2, 13, and 14. Scharff and colleagues\textsuperscript{62} also administered 1 and 2 mg of eszopiclone or placebo nightly to 231 older adults for two weeks, employing nightly patient-reported data.

Erman and colleagues\textsuperscript{60} evaluated multiple dosages of eszopiclone (1, 2, 2.5, and 3 mg vs. placebo and an active control (zolpidem 10 mg) in 65 adult subjects (age 21-65) who received each intervention for two nights, followed by 3-7 day washout, in randomized sequences. PSG was conducted for the two nights on each treatment. Primary endpoint was latency to persistent sleep with secondary endpoints of SE and WASO. The Uchimura and colleagues\textsuperscript{63} study employed a similar crossover design with eszopiclone doses of 1, 2, and 3 mg, zolpidem 10 mg and placebo in 65 patients. PSG was conducted during each two-night intervention. Primary endpoints were objective latency to persistent sleep (LPS) and subjective SL. Zammitt and colleagues\textsuperscript{64} examined eszopiclone 2 and 3 mg vs. placebo for 44 consecutive nights, with PSG on nights 1, 15, 29. Patient-reported data were collected for nights 1, 15, 29, 43, and 44. Primary endpoint was PSG-defined LPS.

Krystal and colleagues\textsuperscript{66} investigated six-month nightly use of eszopiclone 3 mg vs. placebo in 788 adults. Patient-reported data were collected at weekly intervals. Similarly, Walsh and colleagues\textsuperscript{67} reported on nightly use of eszopiclone 3 mg in 830 adults, with weekly patient-reported data. Finally, Boyle and colleagues\textsuperscript{65}, in a study designed primarily to assess next-day driving skill, report subjective data from a single night of eszopiclone 3mg vs placebo.

Sleep latency
Three studies assessed latency to persistent sleep (LPS) as determined by PSG for eszopiclone 2 mg.\textsuperscript{61, 63, 64} The McCall investigation\textsuperscript{61} focused exclusively on older adults and, of note, demonstrated the greatest reduction in LPS. The mean reduction in LPS vs. placebo for the three studies (-14.87 min; CI: -5.47 to -24.27 min) exceeds the threshold for clinical significance. The quality of evidence is LOW due to imprecision and potential publication bias.

All six trials of eszopiclone 2 mg reported subjective sleep latency.\textsuperscript{59-64} As noted above, three of the six included only older adults. Mean difference from placebo fell slightly below the clinical significance threshold (-17.78 min; CI: -7.04 to -28.52 min). The quality of this evidence was LOW due to imprecision and potential publication bias.
Three studies investigated PSG LPS with eszopiclone 3 mg.\textsuperscript{63-65} The mean difference in LPS (-13.63 min; CI: -3.7 to -23.56 min) fell below the clinical significance threshold. The quality of evidence was VERY LOW due to heterogeneity, imprecision and potential publication bias. Subjective SL with eszopiclone 3 mg was reported in four studies.\textsuperscript{60, 64, 66, 67} The mean difference exceeded the clinical significance threshold (-25.00 min; CI: -13.94 to -36.07 min). The greatest reductions were reported in the extended 6-month trials of Krystal and Walsh. Quality of evidence was LOW due to imprecision and potential publication bias.

Two additional studies which were not included in the meta-analysis reported subjective SL with eszopiclone 3 mg. Soares and colleagues\textsuperscript{68} analyzed efficacy in perimenopausal/early menopausal women with sleep initiation complaints. Joffe et al.\textsuperscript{69} examined outcomes in perimenopausal/menopausal women who also exhibited hot flashes and manifested either sleep initiation or maintenance problems. The reductions in sleep latency vs. placebo for these two studies (-15.7 and -17.8 min, respectively) were within the overall range found in the meta-analysis.

**Total sleep time**

Only one eszopiclone study reported adequate objective total sleep time data; therefore meta-analysis was not possible for this outcome at either dosage.\textsuperscript{61} Four studies included subjective TST for eszopiclone 2mg.\textsuperscript{60-62, 64} The meta-analysis reveals a mean increase in TST of 27.53 min vs. placebo. This falls just below the threshold for clinical significance for subjective TST of 30 min. The quality of evidence is LOW due to imprecision and potential publication bias. The one study, noted above, which reported objective TST (in patients >65 years) found an increase in TST of 28.6 min greater than placebo, consistent with the subjective results.

Four studies included adequate data for subjective TST meta-analysis for eszopiclone 3 mg.\textsuperscript{60, 64, 66, 67} These studies demonstrate substantially greater increases in TST at this dosage with a mean difference (vs. placebo) of 57.1 min, well in excess of the clinical significance threshold and substantially greater than the increase observed with 2 mg. The quality of evidence was MODERATE, due to potential publication bias.

The two studies of eszopiclone 3 mg in perimenopausal/early menopausal women revealed mean increases in subjective TST (vs. placebo) of +66.5 min and +23.0 min.\textsuperscript{68, 69}

**Wake after sleep onset**

Two studies were included in the meta-analysis of objective WASO for eszopiclone 2 mg.\textsuperscript{61, 64} The mean difference in WASO was 10.02 min (in favor of eszopiclone), well below the clinical significance level of 20 min for PSG data. The quality of evidence was rated as MODERATE due to potential publication bias. The confidence interval (-2.77 to -17.27 min) fell entirely below the threshold.

Five studies reported adequate data for subjective WASO meta-analysis.\textsuperscript{59-62, 64} Mean difference was far below the threshold for clinical significance (-4.74 min; CI -11.87 to +2.39 min). There was a 30 min reduction for subjective WASO. The quality of evidence was MODERATE due to potential publication bias.

The data for PSG and patient-reported WASO with eszopiclone 3 mg demonstrate somewhat greater reduction of WASO than seen with 2 mg but were still below clinical significance levels. The two studies including PSG WASO demonstrated a mean reduction of 14.69 minutes vs. placebo (CI: -11.69 to -17.68 min).\textsuperscript{64, 65} Quality of evidence was MODERATE (potential publication bias). Subjective WASO for 3 mg was reported in four studies with mean reduction of 15.14 minutes (CI: -8.16 to -22.11 min). Quality of evidence was LOW due to imprecision and potential publication bias.

Krystal and colleagues\textsuperscript{66} published an independent sub-group analysis of the subjective WASO data from their 6-month nightly trial of 3 mg for the purpose of evaluating the impact of baseline WASO severity on
outcome. They identified a positive relationship between baseline WASO severity and degree of improvement in WASO (as determined by eszopiclone/placebo difference) at all time points. They note that studies which do not include a minimum WASO in the inclusion criteria may often be underpowered to identify significant improvement in WASO. The two investigations of menopausal women found eszopiclone-placebo mean differences for subjective WASO of 37.3 and 14.9 min, respectively.68,69

Quality of sleep
The meta-analysis for sleep quality with eszopiclone 2 mg included four studies and found a moderate effect size of +0.47 SMD (CI: +0.32 to +0.63 SMD).60,62-64 The quality of evidence was MODERATE due to imprecision and potential publication bias. Sleep quality ratings for 3 mg, based on six studies showed a large effect size of +0.82 SMD (CI: +0.41 to +1.24 SMD), although quality of evidence was VERY LOW due to imprecision, heterogeneity and potential publication bias.60,63-67 In addition to the studies included in meta-analysis, Soares and colleagues68 reported statistically significant improvement in quality for eszopiclone 3 mg in their study of peri- and postmenopausal women.

Sleep efficiency
Two studies reported PSG SE for eszopiclone 2 mg.61,64 Analysis yielded a mean improvement in SE of 4.83%, just below the significance threshold of 5%. The CI crosses the threshold (+2.21 to +7.46%). For the 3 mg dosage, PSG SE exceeded the clinical significance threshold at 5.61%.64,65 The quality of evidence for both doses was LOW due to imprecision and potential publication bias.

In studies outside the meta-analysis, Joffe69 reported a 14.6% improvement vs. placebo in SE with 3 mg.

Number of awakenings
The PSG NOA for 2 mg showed an insignificant increase of 0.12 awakenings based on two studies.61,64 Evidence quality was MODERATE. Subjective NOA was based on four studies and likewise demonstrated no clinically significant difference from placebo. Evidence quality was MODERATE due to potential publication bias.60-62,64

Overall Quality of Evidence
The overall quality of evidence in the meta-analytic data from these studies was downgraded to VERY LOW for several reasons. Substantial heterogeneity across studies was noted for multiple outcomes. The data were also downgraded for imprecision, due to the relatively large confidence intervals, which cross the clinical significance thresholds for several outcomes. All of these studies were industry sponsored, resulting in further downgrading of evidence due to potential publication bias. The quality of evidence for individual outcomes ranged from MODERATE to VERY LOW, therefore the overall quality of evidence was VERY LOW.

Harms
Sufficient data for meta-analysis of side effects was available only for the 2 mg eszopiclone dosage. Five side effects (dizziness, dry mouth, headache, somnolence and unpleasant taste) were included. Four studies examined dizziness with 2 mg eszopiclone and found no difference from placebo.60,61,63,64 Two studies reported adequate data for dry mouth.61,64 A +0.06 risk difference was reported for eszopiclone. For headache, four studies found essentially no difference between eszopiclone and placebo.59,60,62,64 The same was true for next-day somnolence, based on five studies.60-64 Finally, five studies found a +0.07 risk difference for unpleasant taste.59-62,64

Although meta-analysis was not possible for eszopiclone 3 mg, individual studies reported results which are consistent with those of the 2 mg dosage. Krystal and colleagues66 reported somewhat higher adverse event rates for somnolence (eszopiclone 9.1%; placebo 2.6%); unpleasant taste (26.1% vs. 5.6%); dry
mouth (6.6% vs. 1.5%); and dizziness (9.8% vs. 3.1%). Boyle studied braking reaction time and other performance measures and found no difference between eszopiclone 3mg and placebo. Walsh reported lower work limitation scores for the eszopiclone group, as well as greater improvement on various SF-36 scales at various time points in their study. They also found no difference on Benzodiazepine Withdrawal Scale scores following discontinuation. Zammitt demonstrated no impairment in digit symbol substitution at either 2 mg or 3 mg. Joffe and colleagues reported a 15.2% risk for metallic taste, but placebo rate for this side effect was not identified. Soares and colleagues found a significant increase in unpleasant taste with eszopiclone (17.6% vs. 0.5%). Headache frequency was no different and report of dry mouth was slightly increased for eszopiclone (4.0% to 1.4%).

In summary, the task force found that there was weak evidence of efficacy in the treatment of sleep initiation and maintenance insomnia, with limited or no consistent evidence of adverse events in excess of placebo, with the possible exception of unpleasant taste. Therefore, benefits were deemed to marginally outweigh harms.

**Patients’ Values and Preferences**

Based on its clinical judgement, the task force determined that in light of its efficacy for sleep initiation and maintenance, and its relatively benign side effect profile, a majority of patients would be likely to use eszopiclone compared to no treatment.

### 4.3 Zaleplon for the treatment of Primary Insomnia

**4.3a. We suggest that clinicians use zaleplon as a treatment for sleep initiation insomnia (versus no treatment) in adults. [WEAK FOR]**

**Remarks:** This recommendation is based on Zaleplon 5mg and 10mg.

**4.3b Summary**

Two RCTs meeting inclusion criteria investigated the use of zaleplon 5 or 10 mg in the treatment of chronic primary insomnia. One of these reported only subjective outcomes, and one reported subjective and PSG outcomes. No meta-analysis was possible for these studies, due to the manner of reporting results. The overall quality of evidence from these studies was downgraded to LOW due to imprecision and potential publication bias; all reported studies were industry supported. The overall evidence for zaleplon 10 mg support its efficacy for the treatment of sleep initiation insomnia. At the 10 mg dose, one objective (PSG) study demonstrated a reduction in sleep latency from baseline that met our criteria for clinical significance; the approximately 9.5 minute difference from placebo was determined to be clinically significant. Subjective sleep latency, reported in one study, showed non-significant changes in subjective sleep latency of -11.4 minutes. Subjective TST increased by approximately 21.5 minutes, but the difference from placebo was not significant. WASO was not significantly different from placebo. Similarly, subjective sleep quality showed minimal differences from placebo. The overall evidence for zaleplon 5 mg did not support its efficacy for treatment of any insomnia symptoms, based on self-report studies only. No PSG studies at the 5 mg dose met inclusion criteria. Treatment-emergent adverse events showed no significant difference from placebo for zaleplon 10 mg or 5 mg, and only one study suggested a small increase in rebound using self-reported TST as the outcome.

Data from three additional studies of zaleplon 5–10 mg met our inclusion criteria but could not be included in meta-analyses because key outcome data were presented in insufficient detail. However, the results of these three studies were consistent with those of the two studies presented above, in finding differences from placebo in subjective SOL, but no significant differences in subjective TST or sleep quality.
Overall, the evidence for efficacy of zaleplon 10 mg is marginal, and the evidence for harm appears equivalent to placebo, therefore potential benefits minimally outweigh potential harms. The lack of evidence for efficacy of zaleplon 5 mg makes any potential benefits equivalent to its minimal potential harms.

Based on clinical judgment, the task force determined that the majority of well-informed patients would use zaleplon 10 mg over no treatment. This judgement is based on the minimal evidence of improved sleep latency across PSG and self-report domains, coupled with a low potential for adverse events.

See Supplemental Materials; Table S4-S5.

4.3c Discussion
Evidence from three RCTs which investigated the use of zaleplon 5 or 10 mg in the treatment of chronic primary insomnia was included in the main analysis of outcomes, although meta-analysis could not be performed because data were presented as medians, or as means with no standard deviation. Subjects in each study met criteria for primary insomnia or insomnia associated with nonpsychotic mental disorder by either DSM-III-R or DSM-IV criteria, together with quantitative criteria for self-reported sleep disturbance (SOL ≥30 min, plus either subjective TST≤6.5 hours, WASO ≥30 minutes, or ≥3 awakenings) and associated daytime complaints. Walsh 2000 also required PSG LPS of >20 minutes on two screening nights. Patients were 18-65 years of age or 65 years and older. Study designs included randomized, double-blind, placebo run-in with zaleplon 5-20 mg or placebo for 14-35 nights, followed by a 2-7 night placebo substitution. Walsh 2000 used PSG outcomes, whereas the other two studies used self-report only. Data for zaleplon 20 mg were not considered here.

Sleep latency
One study evaluated the impact of zaleplon 10 mg vs. placebo on PSG sleep latency (SL). This study showed a clinically significant 9.5 minute reduction in mean sleep latency vs. placebo (difference in median of 8.5 minutes) that approached the 10 min value considered clinically significant. The confidence interval (-0.19 to -18.80 min) crossed the clinical significance threshold, and therefore the quality of evidence was downgraded for imprecision. It was downgraded further due to the risk of publication bias since the study was industry-funded. The resultant quality of evidence is LOW.

Self-reported sleep latency was reported in one study, which showed a reduction compared to placebo at the end of treatment (-11.40 min; CI: -26.36 to +4.56 min), but failed to meet clinical significance. Hedner also reported reductions in subjective sleep latency; however, the results could not be subject to meta-analysis, since the mean values were presented only in graphic form.

Additional studies not included in the meta-analysis yield similar findings. Ancoli-Israel conducted a randomized, double-blind, multi-center study of the efficacy of zaleplon 5 and 10 mg vs. placebo in older adults with DSM-IV insomnia, using a similar study design to Hedner, with self-report outcomes. This study reported significant differences between zaleplon 10 mg and placebo at both treatment weeks, and between zaleplon 5 mg and placebo at week 2 only. Elie 1999 reported significant differences on placebo at weeks 1 – 3 of treatment, with differences in the range of -8 to -15 minutes. Fry reported a 28-day double-blind, placebo run-in and run-out study of adults with DSM-III-R insomnia. Median subjective sleep latency was significantly different from placebo at weeks 1, 3, and 4 for zaleplon 10 mg, and at week 1 for zaleplon 5 mg. Because mean and standard deviation data were not reported, data from these two studies could not be formally evaluated in our meta-analysis.
**Total sleep time**

The effects of zaleplon 10 mg on subjective TST were evaluated in one study. Over the course of a five-week study, TST differed significantly from placebo only in week one, with a difference of 21.5 minutes between groups (CI: -5.6 to +48.6 min), however this failed to meet clinical significance. Quality of the evidence was downgraded to LOW due to imprecision and potential publication bias.

Objective TST was evaluated in 2 studies; however, meta-analysis of these studies was not possible due to the manner of data reporting. These studies showed no consistent evidence of a zaleplon-placebo difference at the 10 mg or 5 mg dose of zaleplon. Mean/median differences in subjective TST at the end of treatment were in the range of +7 to +22.4 minutes in favor of zaleplon. The results of studies not included in our formal analysis showed very similar findings for subjective TST, with inconsistent differences between placebo and zaleplon 10 mg.

The effects of zaleplon 5 mg versus placebo on subjective total sleep time were reported in one study. No significant differences in median sleep time were found between zaleplon 5 mg and placebo across 2-4 weeks of treatment. The results of studies not included in our formal analysis showed similar findings for subjective TST, with no differences between placebo and zaleplon 5 mg.

**Wake after sleep onset**

Objective WASO was evaluated in one study, but failed to meet clinical significance (-2.10 min; CI: -10.23 to +6.03 min). The quality of evidence was MODERATE, due to potential publication bias. Subjective WASO was not reported in any of the studies.

**Quality of sleep**

Subjective sleep quality, evaluated on an ordinal 1–7 scale (1= good, 7 = bad) was reported in one of the formally evaluated studies for both 5 mg and 10 mg. At both dosages sleep quality improved (-0.10 points; CI: -0.27 to +0.07 points), but failed to meet clinical significance. The quality of evidence for both doses was downgraded to MODERATE due to potential publication bias.

In three additional studies, subjective sleep quality differed from placebo inconsistently at either dose; the majority of study weeks showed no difference between groups. Quality of evidence was downgraded for publication bias. Precision and heterogeneity could not be formally evaluated.

**Sleep efficiency**

Neither PSG nor subjective sleep efficiency were formally evaluated in any of the studies reviewed here.

**Number of awakenings**

Number of awakenings were evaluated in the sole PSG study; however, it could not be included in the formal evaluation. No data were presented in the paper, but NOA was reported not to differ between zaleplon 10 mg and placebo at any treatment week. Subjective NOA was evaluated in the two studies formally included in our evaluation but data were presented as median values and could not be included in meta-analyses. Hedner reported a difference of uncertain clinical significance only at week 1 and Walsh reported a difference only at week 3. Data from three additional studies not included in our formal analysis showed no significant differences in NOA for either zaleplon 10 mg or zaleplon 5 mg at any study week.

**Overall Quality of Evidence**

As noted above, no meta-analyses could be conducted on data from studies of zaleplon. This was due to the fact that some studies reported median data only, or mean values with no standard deviation, for some of the key outcomes. Still other studies presented data for key outcomes only in graphical form. The quality of
evidence was downgraded for imprecision, due to the relatively large confidence intervals which cross the clinical significance thresholds for multiple outcomes. All of these studies were industry sponsored, resulting in further downgrading of evidence due to potential publication bias. The quality of evidence for individual outcomes ranged from moderate to low, therefore the overall quality of evidence was LOW.

**Harms**

No meta-analysis was conducted on harms. Each of the individual studies showed no significant difference in the overall rate of treatment-emergent adverse events (TEAEs) between zaleplon and placebo. Several symptoms related to the central nervous system were more frequent numerically among zaleplon treated patients, although these differences were not statistically significant due to the low overall incidence of adverse events. The most common TEAEs in studies of zaleplon versus placebo included headache, asthenia, neurasthenia, pain, fatigue, and somnolence. There was no clear evidence of dose-dependent effects.

Several of the reviewed studies reported data from double-blind placebo runout periods. No significant withdrawal symptoms were noted on the Benzodiazepine Withdrawal Symptom Questionnaire. The single PSG study noted no evidence of withdrawal upon discontinuation for the 10 mg dose. Evidence of discontinuation-related increases in subjective TST were noted at the zaleplon 5 and 10 mg dose in older adults, and for subjective SOL in older adults at the zaleplon 5 mg dose. A small increase in NOA of the second discontinuation night was also noted with zaleplon 5 mg. These differences were small in absolute magnitude and of doubtful clinical significance. Other studies did not find evidence of rebound insomnia (Fry, 2000) Categorically defined rebound insomnia was not significantly different for zaleplon 5 mg or zaleplon 10 mg versus placebo.

The task force found that there was weak objective evidence of efficacy for zaleplon 10 mg in the treatment of sleep initiation insomnia that was just below criteria for clinical significance, and no consistent evidence for efficacy in TST. Likewise, there was no statistical evidence of adverse events in excess of placebo, although some TEAEs were more prevalent in zaleplon groups. Evidence for withdrawal effects was weak, inconsistent, and unlikely to be clinically important. On balance, benefits were deemed to marginally outweigh harms.

**Patients’ Values and Preferences**

Based on its clinical judgement, the task force determined that in light of its efficacy for sleep initiation and its relatively benign side effect profile, a majority of patients would likely use zaleplon compared to no treatment.

See **Supplemental Materials; Tables S4-S5.**

**4.4 Zolpidem for the Treatment of Primary Insomnia**

**4.4a** We suggest that clinicians use zolpidem as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. [WEAK FOR]

Remarks: This recommendation is based on zolpidem 10 mg*.

**4.4b Summary**

Twelve RCTs evaluated zolpidem 10 mg for the treatment of chronic primary insomnia. The overall quality of evidence was downgraded to VERY LOW due to significant heterogeneity, imprecision and risk of publication bias. The evidence for zolpidem 10 mg was weakly in favor of its effectiveness for improving sleep initiation, sleep maintenance, sleep quality, SE and TST. In addition, one paper evaluated
the effectiveness of zolpidem extended release 6.25 mg\textsuperscript{83} and one paper assessed zolpidem extended release 12.5 mg\textsuperscript{84}.

Five studies examined the effects of zolpidem 10 mg on objective sleep latency.\textsuperscript{63, 76, 79, 80, 85} The mean reduction (versus placebo) for PSG-determined latency to sleep exceeded the threshold for clinical significance. Ten studies presented patient-reported sleep latency data.\textsuperscript{60, 63, 73, 75-79, 81, 85} The mean reduction in subjective latency fell approximately at the significance threshold. Two studies\textsuperscript{76, 79} reported adequate objective TST data for meta-analysis and found that the mean improvement in TST also exceeded the clinical significance threshold. The same was true for subjective TST, based on eight studies.\textsuperscript{60, 73, 76-79, 81, 85} Two studies\textsuperscript{76, 79} found that PSG-determined WASO reduction was clinically significant. Six studies included adequate data for meta-analysis of subjective WASO\textsuperscript{60, 75, 78, 79, 81, 85}; mean reduction fell substantially below the clinical significance threshold. Six studies evaluating sleep quality reported moderately large improvement in this parameter based on standard mean difference.\textsuperscript{60, 63, 79, 81, 82, 85} Improvement in PSG SE in the four studies included also exceeded the significance threshold.\textsuperscript{76, 79, 80, 85} Number of awakenings (objective) fell below the clinical significance threshold.\textsuperscript{60, 85} Reduction in subjective number of awakenings also failed to meet the significance threshold.

The single paper reporting on extended-release zolpidem 6.25 mg\textsuperscript{83} found moderate reduction in polysomnographically-determined WASO (based on only the first 6 hr of sleep) and minimal improvement in LPS and SE at end-treatment in an elderly population. Overall quality of evidence from this report was LOW due to imprecision and potential publication bias. Data from the one study\textsuperscript{84} on zolpidem extended-release 12.5 mg found moderate reduction in PSG LPS. Reduction in WASO was also moderate, while SE was not significantly different from placebo. Overall quality of evidence was LOW due to imprecision and potential publication bias.

Meta-analysis was conducted for amnesia, dizziness, headache, nausea, somnolence and taste perversion in studies employing zolpidem 10 mg. Small, but potentially significant increases in amnesia, dizziness and somnolence were reported with zolpidem.

Overall, the benefits of zolpidem 10 mg and extended-release zolpidem 12.5 mg were judged to be greater than the minimal potential harms. Benefits and harms were judged to be approximately equal for extended-release zolpidem 6.25 mg. It was determined by clinical judgement of the task force that the majority of well-informed patients would use zolpidem 10 mg and extended-release 12.5 mg over no treatment. This judgement is based on the evidence of improvement in sleep latency, total sleep time, WASO, sleep efficiency and sleep quality, coupled with relatively low potential for adverse events. The data for efficacy of zolpidem extended-release 6.25 mg is minimal and inconclusive at best.

* See summary regarding extended-release zolpidem

See **Supplemental Materials**, Figures S26-S35 and S72-S77, and Tables S6-S8.

### 4.4c Discussion

A total of twelve studies were included in the meta-analysis for zolpidem 10 mg.\textsuperscript{60, 63, 73, 75-82, 85} Dorsey and colleagues\textsuperscript{75} studied 141 menopausal or perimenopausal women who exhibited both insomnia (TST <6hr or WASO >1hr) and nocturnal hot flashes or sweats. Subjects received zolpidem 10 mg or placebo in a 4 week trial. Outcomes included patient-reported TST, SL, WASO, and NOA. Elie\textsuperscript{73} investigated three dosages of zaleplon versus zolpidem 10 mg or placebo. The study included 615 adults with SL >30 min and either TST <6.5 hr or WASO >30 min or >3 awakenings per night. Subjects received one of three zaleplon dosages, zolpidem 10 mg or placebo for 28 nights. Outcome data included subjective SL, quality of sleep (QOS), TST and NOA. Erman\textsuperscript{60} assessed 65 adults with reported sleep-onset insomnia and baseline PSG SL >20 min
and TST <7 hr or WASO >20 min. Enrollees were administered eszopiclone at 4 dosages, zolpidem 10 mg and placebo in a randomized treatment sequence of 2 nights per intervention with intervening washout. Primary outcome was PSG-determined LPS with secondary measures including SE, WASO and NOA. Hermann administered zolpidem 10 mg or placebo for two weeks to 21 adults with difficulty initiating or maintaining sleep. PSG was conducted on the final treatment night with reported outcomes including SL, TST, SE and WASO.

Perlís evaluated 199 subjects with primary insomnia (SL ≥45 min or TST ≤6 hr) with zolpidem 10 mg or placebo. Subjects were instructed to take the medication 3-5 times per week as needed over a twelve week period. Sleep diary outcomes included SL, TST, WASO and NOA. Jacobs compared zolpidem 10 mg, cognitive behavior therapy and placebo in 63 adults with primary sleep-onset insomnia (SL >1 hr on ≥3 nights/week). Subjects received zolpidem for 28 days, followed by taper. Primary outcome was patient-reported sleep latency with secondary outcomes of SE and TST. Randall investigated the efficacy of zolpidem 10 mg (5 mg for subjects 65-70 years) over an eight month period in 91 subjects (age 23-70 years) with screening PSG SE <85%. Patient-reported outcomes and PSG data at one and eight months included SL, TST, WASO and SE. Schär evaluated 75 adults for five weeks with zolpidem 10 mg, 15 mg or placebo. Inclusion criteria included SL ≥30 min or TST <6 hr. Subjects underwent sleep studies on the first two nights of each treatment week. Primary outcomes were defined as LPS and SE.

Staner assessed the effects of three drugs, including zolpidem 10 mg, in a driving simulation study of 23 adults with recurrent SL >30 min or WASO >60 min. Sleep quality data was reported. Uchimura compared zolpidem, eszopiclone and placebo in a crossover design as described in the eszopiclone section. Walsh compared zolpidem 10 mg to trazodone 50 mg and placebo in 278 adults with insomnia characterized by frequent SL >30 min and TST 4-6 hr. Subjective sleep latency and TST were reported. Ware assessed rebound insomnia in zolpidem 10 mg, triazolam 0.5 mg and placebo. Ninety-nine subjects with baseline PSG-determined LPS >20 min and TST 4-7 hr took zolpidem 10 mg, triazolam or placebo for 28 consecutive days. PSG LPS, SE, TST, and WASO were evaluated.

Two studies reported on extended-release (ER) zolpidem. Roth assessed zolpidem ER 12.5 mg in 212 adults with insomnia who reported >1 hr WASO at least 3 nights per week. Patients received zolpidem or placebo nightly for 3 weeks in a parallel group design. Walsh studied 205 elderly adults with insomnia with the same inclusion criteria and design, employing a 6.25 mg dose of zolpidem ER versus placebo.

Fourteen additional studies met inclusion criteria but could not be included in meta-analysis due to inadequate data sets. Pertinent results from these studies are noted independently of meta-analysis results.

**Sleep latency**

Five studies included adequate data for PSG SL meta-analysis. The mean difference from placebo of -11.65 min exceeded the clinical significance threshold. The 95% confidence interval of -4.15 to -19.15 min crossed the clinical significance threshold and was therefore considered imprecise. Heterogeneity was high. With potential publication bias as well, the quality of evidence was rated as VERY LOW.

Ten of the twelve studies used in meta-analysis reported subjective SL. The improvement in sleep latency versus placebo was approximately at the significance threshold (mean difference: 19.55 min; CI: -14.2 to -24.9 min). Evidence quality was VERY LOW due to imprecision, heterogeneity and potential publication bias.

Six additional studies assessed sleep latency outcomes with zolpidem. These studies varied significantly with regard to drug preparation, dosage, mode of administration and methodology, rendering
comparisons between them or to the meta-analytic data difficult. Four of the six evaluated sublingual zolpidem, primarily for treatment of middle-of-the-night (MOTN) awakenings. Roth and colleagues\textsuperscript{91} reported results of a three-way crossover study of zolpidem sublingual 1.75 mg, 3.5 mg and placebo. Zolpidem reduced both objective (latency to persistent sleep) and subjective latency to sleep (SL) following MOTN awakenings (PSG: 1.75 mg: -11.2 min vs. placebo; 3.5 mg: -18.4 min vs. placebo; 3.5 mg: -15.2 min). Roth\textsuperscript{92} also reported reduced subjective latencies following MOTN awakenings with sublingual zolpidem 3.5 mg over a 28-day trial. Zammit\textsuperscript{98} administered immediate release oral zolpidem 10 mg, zaleplon 10 mg or placebo to subjects with sleep maintenance insomnia following induced MOTN awakenings. Zolpidem reduced PSG latency to persistent sleep following the awakenings (-30.5 min vs. placebo). Staner\textsuperscript{93} compared the effects of sublingual zolpidem 10 mg to immediate release oral zolpidem on PSG initial sleep latency and reported shorter latency to persistent sleep with the sublingual preparation (-10.28 min) versus the oral preparation. Walsh\textsuperscript{97} investigated subjective SL in an 8-week trial of as-needed zolpidem 10 mg (3-5 times per week). For medication nights only, end treatment SL for the zolpidem 10 mg group was 12.6 min less than the placebo group.

Walsh\textsuperscript{83} investigated the effects zolpidem ER 6.25 mg and found reduction of PSG LPS of 13.0 min. Roth\textsuperscript{84} reported a decrease in PSG LPS of 8.2 min versus placebo at end of treatment.

**Total sleep time**

Two studies\textsuperscript{76, 79} were included in the meta-analysis of polysomnographically-determined TST. Mean reduction in TST met the clinical significance threshold at +28.91 min, however the 95% confidence interval crossed the threshold (CI: +10.85 to +46.97 min). The quality of evidence was downgraded to LOW due to imprecision and potential publication bias. Eight studies reported adequate data for meta-analysis of patient-reported TST.\textsuperscript{60, 73, 76-79, 81, 85} The mean difference for subjective SL from these studies slightly exceeded the significance threshold (+30.04 min; CI: +15.12 to +44.96 min). Quality of evidence was LOW due to imprecision and potential publication bias.

Six additional studies presented TST data which was not sufficient to be included in the analysis.\textsuperscript{74, 86, 87, 91, 95, 98} Allain and colleagues\textsuperscript{86} evaluated zolpidem 10 mg administered on an as-needed basis over a four week period. When only drug nights were included in analysis, zolpidem produced a statistically significantly greater increase in subjective TST versus placebo (+19.9 min). Cluydts\textsuperscript{87} and Hajak\textsuperscript{98} found no difference in subjective TST with nightly versus intermittent (5/7 nights) use of zolpidem 10 mg, both of which produced significant improvement. In a study designed primarily to address potential rebound insomnia following four weeks of treatment with zaleplon, zolpidem or placebo, Fry\textsuperscript{74} reported substantially greater improvement in patient-reported TST with zolpidem 10 mg versus placebo (+28.2 min). In another study of rebound insomnia, Voshaar\textsuperscript{95} compared zolpidem 10 mg to temazepam 20 mg (without placebo control) administered nightly for four weeks. The two drugs produced improved TST without significant difference between the two. Finally, two studies Roth\textsuperscript{92} and Zammit\textsuperscript{98} investigated effects of zolpidem following MOTN awakenings. Roth and colleagues compared sublingual zolpidem 1.75 mg and 3.5 mg to placebo. Both dosages produced greater TST following awakening as compared to placebo (+14.7 min and +25.9 min, respectively). Zammit administered zolpidem 10 mg following MOTN awakening and reported TST after awakening 30 min greater than placebo.

TST data were not reported in the extended-release studies.

**Wake after sleep onset**

Two studies reported adequate data for meta-analysis of PSG WASO.\textsuperscript{76, 79} These studies yielded a mean difference from placebo of -25.46 min (CI: -17.94 to -32.99 min). This exceeds the threshold for clinical significance. The quality of evidence was LOW due to imprecision and potential publication bias.
Zolpidem ER 12.5 mg produced WASO of 20 min less than placebo at treatment conclusion, although this is based on only the first 6 hr of sleep.\textsuperscript{84} Comparison of changes from baseline in this study, however, suggests smaller differences between drug and placebo. Walsh, using the same selective sample of 6 hr, found WASO 13.0 min less in the zolpidem ER 6.25 mg group than in the placebo group. Given the sampling of only 6 hr, it is difficult to clearly determine whether or not these agents would fulfill the criterion for clinical significance, which is based on an entire night of study.

Six studies assessed patient-reported WASO.\textsuperscript{60, 75, 78, 79, 81, 85} The mean difference fell well below the level of clinical significance at 113.57 min (CI: -7.30 to -19.84 min). Quality of evidence was LOW due to heterogeneity and potential publication bias.

\textbf{Quality of sleep}

Six studies included sleep quality data.\textsuperscript{60, 63, 79, 81, 82, 85} The meta-analysis produced a standardized mean difference of +0.64 (CI: +0.03 to +1.26 SMD), suggesting moderate overall improvement in patient-reported sleep quality. Quality of evidence was VERY LOW due to imprecision, heterogeneity and potential publication bias.

\textbf{Sleep efficiency}

Sleep efficiency was reported in four studies.\textsuperscript{76, 79, 80, 85} The mean difference favored zolpidem (+6.12%; CI: +4.39 to +7.85%). Quality of evidence was LOW.

In the Roth\textsuperscript{84} study of zolpidem ER 12.5 mg, PSG SE was 3.9% better with zolpidem than placebo. Walsh found a difference of 2.4% on favor of zolpidem ER 6.25 mg.

\textbf{Number of awakenings}

PSG-determined frequency of awakenings was reported by Schar\textsuperscript{85} and Ware\textsuperscript{80}. The mean difference from placebo was -0.95 awakenings (CI: -0.49 to -1.41 awakenings). This fails to meet the clinical significance threshold. Quality of evidence was MODERATE. Subjective awakening was reported in six studies.\textsuperscript{73, 75, 76, 78, 81, 85} Mean reduction versus placebo was -0.31 awakenings (CI: -0.17 to -0.45 awakenings). This also fails to achieve clinical significance. Evidence quality was LOW due heterogeneity and potential publication bias.

\textbf{Overall Quality of Evidence}

The overall quality of evidence in the meta-analytic data from these studies was downgraded to VERY LOW for several reasons. Substantial heterogeneity across studies was noted for multiple outcomes. The data were also downgraded for imprecision, due to the relatively large confidence intervals, which cross the clinical significance thresholds for several outcomes. All of these studies were industry sponsored, resulting in further downgrading of evidence due to potential publication bias. The quality of evidence for individual outcomes ranged from MODERATE to VERY LOW, therefore the overall quality of evidence was VERY LOW.

\textbf{Harms}

Meta-analysis for adverse effects of zolpidem was possible for six side effects: amnesia, dizziness, sedation, headache, nausea, and taste perversion. Two studies\textsuperscript{73, 85} included data on amnesia and found a minimal difference from placebo (0.03 risk difference). A small increase in risk (0.06 risk difference) was identified for dizziness, based on analysis of four investigations.\textsuperscript{60, 63, 75, 85} Risk for headache was mildly increased in the zolpidem group (0.07 risk difference).\textsuperscript{60, 75, 81} Minimal difference was observed in the risk for nausea (0.02 risk difference)\textsuperscript{60, 85}, while somnolence was found to have only slightly higher risk (0.04), based on six studies.\textsuperscript{60, 63, 73, 75, 81, 85} Risk for taste perversion was low and approximately equal in both groups.\textsuperscript{63, 73}
Numerous studies have evaluated rebound insomnia after discontinuation of zolpidem,\textsuperscript{71, 73, 74, 76, 78, 80, 83, 85, 89, 95} Many of these studies found no evidence of rebound after varying durations of nightly or intermittent use, for up to six months.\textsuperscript{71, 76} Some investigations reported evidence of rebound, limited primarily to night 1 following discontinuation.\textsuperscript{73, 74, 83, 84}

Evaluation of daytime improvement and impairment was limited. Dorsey\textsuperscript{75} reported improvement in sleep-related difficulty with daytime function. Hajak\textsuperscript{88} described marked improvement in quality of life ratings with both nightly and intermittent use. Morning alertness and performance impairment were tested in several studies. Roth\textsuperscript{84} and Walsh\textsuperscript{83} found no evidence of impairment on digit symbol substitution test (DSST) or Rey auditory-verbal learning test (RAVLT) after zolpidem modified-release 12.5 mg. Scharf\textsuperscript{85} reported no impairment on DSST or digit symbol copying. Staner\textsuperscript{82} found no indication of impairment in a driving simulation study after seven consecutive nights of zolpidem 10 mg. Zammit\textsuperscript{98} formally assessed sleepiness following administration of zolpidem 10 mg following MOTN awakening. Subjects showed significantly reduced PSG latencies versus placebo at 4, 5, and 7 hours following administration. This was accompanied by impairment on DSST at 4 and 5 hours.

In summary, the task force found that there was weak evidence of efficacy in the treatment of sleep initiation and maintenance insomnia, with limited evidence of mild adverse events in excess of placebo, with the possible exception of excessive sleepiness following administration of higher dosages (10 mg) less than 8 hours prior to awakening. Therefore, benefits were deemed to marginally outweigh harms.

**Patients’ Values and Preferences**

Based on its clinical judgement, the task force determined that in light of its efficacy for sleep initiation and maintenance, and its relatively benign side effect profile, a majority of patients would be likely to use zolpidem compared to no treatment.

**Benzodiazepines**

4.5 **Triazolam for the Treatment of Primary Insomnia**

4.5a **We suggest that clinicians use triazolam as a treatment for sleep onset insomnia (versus no treatment) in adults. [WEAK FOR]**

Remarks: This recommendation is based on triazolam 0.25 mg.

4.5b **Summary**

Only one study\textsuperscript{99} contained data sufficient for meta-analysis which was, therefore, not performed. The quality of evidence for this study was HIGH. This study, consisting of patient-reported data, shows a modest decrease in subjective SL, which fell below the clinical significance threshold. Two additional studies, which did not contain data suitable for meta-analysis, did report statistically significant reduction in subjective SL with triazolam 0.25 mg versus placebo.\textsuperscript{100, 101} Roehrs\textsuperscript{99} found an increase in TST, although this fell somewhat below the range of clinical significance as well. WASO was not reported, while sleep quality showed mild to moderate reduction. Number of awakenings was insignificantly decreased.

No meta-analysis of harms was possible. Given the marginal evidence for efficacy in improving sleep onset, coupled with limited evidence regarding harms, the task force judged the harms to be approximately equal to the benefits. Based on its clinical judgement, the task force determined that, in light of the evidence for efficacy for sleep onset and the absence of information regarding harms, the majority of patients would be likely to use triazolam 0.25 mg compared to no treatment.
See Supplemental Materials; Tables S9.

4.5c Discussion

Roehrs\textsuperscript{99} studied 32 adults with insomnia in a complex design which began with 11 days in which subjects received either triazolam or placebo nightly, "as needed" or every third night. This was followed by 14 nights in which subjects chose to self-administer treatment, with placebo (week 1) or triazolam 0.25 (week 2).

Thirteen additional studies met general inclusion and exclusion criteria.\textsuperscript{100-112} These studies were highly varied in design, many utilizing interval scales (as opposed to specific numeric values) for reporting of sleep outcome variables. Some did not include a placebo comparison. Many included dosages which are substantially higher than current recommended dosages. Therefore, only those studies which contained pertinent data are discussed.

Bowen\textsuperscript{103} compared triazolam 0.5 mg, flurazepam 30 mg and placebo in 120 insomnia outpatients. The two-night crossover comparison of triazolam 0.5 mg and placebo included only 18 subjects, who completed morning sleep questionnaires. Elie\textsuperscript{100} evaluated triazolam 0.125 mg (with upward dosage adjustment to 0.25 mg during the study period, as indicated) versus zopiclone and placebo in 48 elderly (60-90 years) subjects. Subjects received one of three interventions nightly for three weeks in a parallel group design. Outcome variables were patient-reported. Greenblatt\textsuperscript{106} reported an RCT of 6 nights baseline placebo administration followed by triazolam 0.5 mg for 7-10 nights in a total of 60 subjects with sleep initiation or maintenance insomnia. Data were subjective reports. Hajak\textsuperscript{107} evaluated 1507 subjects with sleep initiation or maintenance insomnia with triazolam 0.25 mg, zopiclone or placebo. The triazolam vs. placebo comparison groups totaled 605 subjects who received drug or placebo for 28 consecutive nights and reported sleep variables on visual analog scales.

Monti\textsuperscript{109} assessed 24 chronic insomnia subjects with triazolam 0.5 mg, zolpidem and placebo in a 27 night trial, with PSG on nights 4/5 and 15/16 and 29/30. Reeves\textsuperscript{101} evaluated 37 geriatric subjects (> 60 years) with triazolam 0.25 mg, flurazepam or placebo in a 28 day trial. The triazolam and placebo groups included 28 subjects who completed daily sleep diaries. Rickels\textsuperscript{110} studied 50 subjects with sleep initiation or maintenance insomnia who received either triazolam 0.5 mg or placebo for 7 days. Outcome data were subjective ratings and interval scales. Scharf\textsuperscript{111} administered triazolam 0.5 mg, quazepam or placebo to 65 chronic insomnia subjects. After placebo run-in, participants received nightly drug or placebo for 9 nights, followed by 14 nights of every other night administration. Outcomes were patient reported rating scales.

**Sleep latency**

In the only study with adequate data for meta-analysis, Roehrs\textsuperscript{99} found a small reduction in subjective SL (-9.2 min; CI: -22.3 to +3.9 min) which fell below clinical significance. Quality of evidence for these data was HIGH.

Likewise, Elie\textsuperscript{100} found subjective ratings for reduction of SL with triazolam 0.125-0.25 mg superior to placebo. Hajak found no significant difference from placebo in SL “response rate” (SL reduction of ≥15 min) for triazolam 0.25 mg. In contrast, Reeves\textsuperscript{101} found triazolam 0.25 mg statistically superior to placebo for SL in a geriatric population on ratings of "how much [the drug] helped.”
Bowen\textsuperscript{103} noted triazolam 0.5 mg to be significantly better than placebo in interval ratings for reduction of sleep onset time. Greenblatt reported sleep diary reductions from baseline placebo levels of 55 min and 24 min in two separate triazolam 0.5 mg groups. Rickels\textsuperscript{110} reported similar subjective improvement on ratings of sleep induction for triazolam 0.5 mg. Monti\textsuperscript{109} found no significant differences between triazolam 0.5 mg and placebo for PSG SL at any time point.

**Total sleep time**

Roehrs\textsuperscript{99} observed a moderate increase in subjective TST (+25.20 min; CI: -9.12 to +59.52 min). This fell only slightly below the clinical significance threshold of 30 m but was not statistically different from placebo. Quality of evidence was MODERATE due to imprecision.

In additional studies, Hajak\textsuperscript{107} found no significant difference between triazolam 0.25 mg and placebo in “percentage of responders” for TST (defined as >20% increase in TST). Ratings for improvement in TST were significantly better for triazolam 0.25 mg than placebo in the Reeves\textsuperscript{101} geriatric study.

Monti\textsuperscript{109} observed a statistically significant increase in objective TST with triazolam 0.5 mg (+16 min versus placebo). Bowen\textsuperscript{103} found that triazolam 0.5 mg was significantly preferred to placebo for sleep duration. In two separate triazolam 0.5 mg groups, Greenblatt\textsuperscript{106} noted increases in subjective TST of 1.02 and 0.76 hr. Rickels\textsuperscript{110} reported that triazolam 0.5 mg was rated as significantly superior to placebo for sleep duration. Scharf\textsuperscript{111} noted significantly greater subjective improvement in interval ratings of TST with both daily and every other night administration of triazolam 0.5 mg.

**Wake after sleep onset**

No studies reported data on WASO.

**Quality of sleep**

Roehrs\textsuperscript{99}, using a 4-point scale (1-“good”, 4-“poor”), found mild improvement in QOS (-0.37 points; CI: -0.66 to -0.07 points), which was not considered to be clinically significant Quality of evidence was HIGH.

Elie\textsuperscript{100} found no significant difference between triazolam (0.125/0.25 mg) in an elderly population with respect to QOS. Likewise, Hajak\textsuperscript{107} reported no statistically significant difference between triazolam 0.25 and placebo. Reeves\textsuperscript{101} demonstrated significant improvement in QOS versus placebo at the same dosage in a geriatric population.

**Sleep efficiency**

Sleep efficiency was not reported by any study.

**Number of awakenings**

Roehrs\textsuperscript{99} reported a reduction in NOA of 0.37 (CI: -1.7 to +0.96 awakenings). This did not meet the clinical significance threshold. Quality of evidence was LOW due to significant imprecision.

Hajak\textsuperscript{107} noted no significant difference from placebo in the percentage of “responders” (reduction of NOA to ≤3) with triazolam 0.25 mg. However, Reeves\textsuperscript{101} did find a statistically significant reduction at this dosage in ratings for NOAs. Bowen\textsuperscript{103} observed a statistically significant reduction in subjective ratings for NOA with triazolam 0.5 mg versus placebo. Greenblatt\textsuperscript{106} also reported reductions of 0.58 and 0.89 patient-reported awakenings from placebo baseline in two groups administered 0.5 mg triazolam.
**Overall Quality of evidence**
The overall quality of evidence for the triazolam data, based on the single study meeting criteria for meta-analysis, was HIGH.

**Harms**
Insufficient data were available for meta-analysis of adverse events associated with triazolam 0.25 mg. Very little systematic analysis of adverse effects is available. Hajak\textsuperscript{107} reported that “speech disorder” was the only adverse effect, among many, to be significantly more frequent in the triazolam group than in the placebo condition.

**Patients’ Values and Preferences**
Based on its clinical judgement, the task force determined that, in light of the minimal evidence for efficacy and the absence of information regarding harms, the majority of patients would not be likely to use triazolam 0.25 mg compared to no treatment.

4.6 **Temazepam for the treatment of Primary Insomnia**

4.6a **We suggest that clinicians use temazepam as a treatment for sleep initiation and sleep maintenance insomnia (versus no treatment) in adults. [WEAK FOR]**

**Remarks:** This recommendation is based on temazepam 15 mg.

4.6b **Summary**
Three RCTs investigated the use of temazepam in the treatment of chronic primary insomnia.\textsuperscript{113-115} These studies provide a limited assessment of temazepam in that they included small sample sizes of 19, 20, and 34 subjects, respectively. The overall quality of evidence from these studies is MODERATE. Meta-analysis for temazepam 15 mg was conducted for SL, TST and sleep quality. Two studies\textsuperscript{113, 115} were included in the meta-analysis of SL. These showed a reduction in subjective SL which just exceeded the threshold for clinical significance. The quality of evidence for SL data was MODERATE. Meta-analysis of TST showed robust improvement in subjective TST which well exceed the threshold for clinical significance. Quality of evidence was MODERATE. There was insufficient data for meta-analysis of WASO. One study of objective WASO revealed a clinically significant reduction. Subjective and objective SE was significantly increased, based on limited data from secondary studies. There was evidence for marginal improvement in sleep quality of 0.25 standard deviations. This was not a clinically significant difference from placebo and falls below the threshold for clinical significance. There were minimal data on adverse effects and what data is available does not suggest a high frequency of TEAEs.

Meta-analysis for temazepam 30 mg was not possible for any sleep outcomes. Data from individual studies are reported below.

In summary, meta-analysis data are available for temazepam 15 mg only. These data, when coupled with that of secondary studies which did not contain data adequate for meta-analysis, demonstrate efficacy for temazepam 15 mg in improving subjective and possibly objective sleep latency, subjective and objective TST and objective WASO (the latter based on a single study). Temazepam 30 mg appears to have significant efficacy for improving subjective sleep latency and TST. The data also support a clinically significant effect for both 15 mg and 30 mg on subjective NOA, although very limited data for objective NOA (at 20 mg) revealed no significant effect.
Given the significant improvements in patient-reported SL and TST, coupled with additional data derived from secondary studies (see below), the TF judged that the benefits of temazepam 15 mg appear to be greater than the potential harms. It was determined by clinical judgement of the task force that the majority of well-informed patients would use temazepam over no treatment.

4.6c Discussion
Evidence from three RCTs which investigated the use of temazepam in the treatment of insomnia was included in the statistical analysis\(^\text{113-115}\)

Glass\(^\text{113}\) evaluated 19 subjects 70 years of age or older who met DSM-IV diagnostic criteria for primary insomnia. Subjects underwent a cross-over study of two weeks of treatment with placebo, temazepam 15 mg or diphenhydramine 50 mg with randomized order of administration. Sleep was assessed using diary-derived variables. Adverse effects were recorded and daytime impairment was systematically assessed using the digit symbol substitution task and the manual tracking task, and the morning-after memory impairment, using a free-recall procedure.

Wu\(^\text{115}\) assessed 71 patients with DSM-IV diagnosed insomnia who were randomized to one of four groups (CBT alone, CBT plus pharmacotherapy with temazepam 15 mg, pharmacotherapy alone, or placebo). For the purpose of this analysis, pharmacotherapy alone was compared to placebo (N=34). Subjects received 8 weeks of treatment. End-of-treatment PSG and patient diary data for SL, TST and SE are compared.

Hindmarch\(^\text{114}\) studied 20 individual with “a history of nighttime medication for insomnia” No additional diagnostic information was provided. Subjects were randomized to receive temazepam 15 and 30 mg or placebo for a single night using a within-subjects cross-over design. Outcome was assessed using the Leeds Sleep Evaluation Questionnaire which consisted of Visual Analogue Scale (VAS) ratings of “Ease of Falling Asleep” and “Quality of Sleep”. Adverse effects were not reported but daytime sedation was assessed with a Choice Reaction Time task, the Critical Flicker Fusion Test, and “Ease of Awakening” and “Integrity of Behavior Following Wakefulness” items from the Leeds scale.

Six additional studies which included temazepam-placebo comparisons were reviewed.\(^\text{95, 116-120}\) Cuang\(^\text{116}\) studied 60 adult “outpatients with insomnia.” Parallel group design included three groups: temazepam 20 mg, temazepam 10 mg and placebo. Subjects received treatment or placebo for five nights. Patient-reported data including sleep quality (“better, same or worse”), SL, and TST were collected each morning. Fillingim\(^\text{117}\) evaluated 75 adult patients with difficulty initiating (SL >30 min) and maintaining (>1 awakening with difficulty returning to sleep) sleep and TST <6 hr. Subjects received temazepam 30 mg, glutethimide or placebo in parallel group design for four nights. Outcomes included patient-reported estimates of SL, TST, NOA and QOS. Heffron\(^\text{118}\) reported on 55 “insomnia outpatients” who received temazepam 30 mg or placebo in parallel groups for four nights. Subjects reported SL, TST, NOA and QOS. Tuk\(^\text{119}\) studied 21 “primary sleep-onset insomnia” patients in a within-patient cross-over study. Subjects received a single night of placebo and a single night of temazepam 20 mg with one week intervening washout. PSG was conducted on each of the two nights. SL, TST, WASO and SE were reported. Voshaar\(^\text{95}\) assessed 85 individuals with DSM-III-R primary insomnia in a within-subjects crossover design including temazepam 20 mg, zolpidem and placebo. A single-blind placebo period of four days was followed by 28 days of active treatment with zolpidem or temazepam. Data are presented as means for the placebo period and active treatment period for each sleep outcome. Wilson\(^\text{120}\) conducted an actigraphic evaluation of 38 subjects with “complaints of poor sleep.” Subjects were randomized to one of two crossover designs, each of which included two weeks of placebo and two weeks of temazepam 20 mg. Subjective results from patient diaries as well as actigraphic results were averaged over the respective periods.
**Sleep Latency**

The meta-analysis for subjective SL, based on two studies\textsuperscript{113, 115} of temazepam 15 mg revealed a mean reduction of 20.06 min (CI: -1.07 to -39.05 min). Quality of evidence was MODERATE due to imprecision.

One additional study assessed subjective SL at the 15 mg dosage. Hindmarch\textsuperscript{114} found no effect on the VAS scale rating for "ease of getting to sleep."

Three studies\textsuperscript{117, 118} evaluated the effects of temazepam 30 mg on subjective sleep latency from patient diaries. Fillingham\textsuperscript{117} reported a reduction of SL of 40 min versus placebo. Heffron\textsuperscript{118} found a 45 min reduction versus placebo. Hindmarch\textsuperscript{114} reported a statistically significant effect on a VAS for "ease of getting to sleep" with temazepam 30 mg.

Tuk\textsuperscript{119} found no difference between temazepam 20 mg and placebo in PSG SL. However, it is noteworthy that in this sample of “primary sleep onset insomnia” patients, both temazepam and placebo produced a reduction from baseline of approximately 53 min (to SL of about 24 min). Wilson\textsuperscript{120} demonstrated a SL derived from actigraphy which was only 7 min less than that of placebo. However, of note, the end-of-treatment SL for temazepam (by actigraphy) was only 15 min, suggesting a possible floor effect for these results.

Three studies assessed subjective SL with temazepam 20 mg.\textsuperscript{95, 116, 120} Cuanang\textsuperscript{116} reported a reduction from baseline which was 34.2 min greater than placebo reduction. Voshaar\textsuperscript{95} found end-of-treatment SL for temazepam 20 mg which was 29 min less than placebo. Similarly, Wilson\textsuperscript{120} found subjective SL was 23 min less than placebo.

**Total sleep time**

Two studies\textsuperscript{113, 115} were included in the meta-analysis for subjective TST at 15 mg. The analysis revealed a mean increase in TST of 64.4 min (CI: +8.1 to +120.8 min). Quality of evidence was MODERATE due to imprecision. No additional studies evaluated subjective TST at this dosage. Wu reported a PSG TST of 99.1 min greater than placebo for 15 mg.

Two studies\textsuperscript{117, 118} reported subjective TST at the 30 mg dosage. Fillingim demonstrated TST which was 53 min greater than placebo, while Heffron noted a 54.6 min greater TST versus placebo. There were no investigations of objective TST for this dosage.

At the 20 mg dosage, three trials\textsuperscript{95, 116, 120} reported subjective TST. Cuanang\textsuperscript{116} found a 78 min greater TST increase from baseline than placebo. Voshaar\textsuperscript{95} demonstrated a 46 min greater TST than placebo at end-of-treatment. Wilson\textsuperscript{120} also found an 18 min greater TST with temazepam 20 mg than with placebo. One study\textsuperscript{120} assessed objective TST at 20 mg. Actigraphy revealed a 12 min greater TST at this dosage versus placebo.

**Wake after sleep onset**

Meta-analysis for WASO was not possible. One investigation\textsuperscript{119} evaluated PSG WASO at the 20 mg dosage and reported WASO time which was 28.1 min less than placebo. Of note, the subjects in this study were described as exhibiting “sleep onset insomnia.” At the same dosage, subjective WASO was 15 min less than placebo\textsuperscript{95}. This was below the threshold for clinical significance.

**Quality of sleep**

Meta-analysis was conducted for sleep quality ratings from two studies\textsuperscript{113, 114} for temazepam 15 mg. The standardized mean difference was 0.25, below the range for clinical significance. However, it should be
noted that the Hindmarch\textsuperscript{114} study was underpowered to detect all but extremely large effects in that it only included 20 subjects. The quality of evidence was MODERATE due to imprecision.

Two studies found statistically significant improvement in sleep quality ratings for temazepam 30 mg.\textsuperscript{117,118} Cuanang reported statistically significant improvement for temazepam 20 mg on a quality rating comparing “better quality” to “same or worse quality.”

**Sleep efficiency**
Meta-analysis was not achievable for SE at any dosage.

At 15 mg, Wu\textsuperscript{115} found a PSG SE which was 13.3\% greater than placebo (CI: +3.9 to +22.6\%). Subjective SE was +14.1\% versus placebo (CI: +5.8 to +22.3\%). The quality of evidence for both was MODERATE due to imprecision. At 20 mg, Tuk\textsuperscript{119} reported a +5.9\% PSG SE versus placebo.

**Number of Awakenings**
No meta-analysis of NOA was possible. One study\textsuperscript{113} reported data for subjective NOA at the 15 mg dosage (-0.5 awakenings; CI: -1.29 to +0.29 awakenings). Quality of Evidence was MODERATE due to imprecision.

Two studies\textsuperscript{117,118} reported subjective NOA at 30 mg. They found -1.0 and -1.24 awakenings, respectively, compared to placebo.

Tuk\textsuperscript{119} found no significant reduction in PSG NOA at 20 mg. One study\textsuperscript{120} reported data for subjective NOA at 20 mg (-0.2 awakenings compared to placebo).

**Overall Quality of Evidence**
The overall quality of evidence in the meta-analytic data from the two available studies was MODERATE for temazepam 15 mg due to imprecision.

**Harms**
Limited data on adverse effects of temazepam 15 and 30 mg are available. Meta-analysis could not be performed. Glass\textsuperscript{113} found no notable increase in adverse effects with temazepam 15 mg vs placebo and no significant effects were found on measures of daytime impairment. Cuanang\textsuperscript{116} reported “no marked difference in adverse events,” although temazepam 20 mg was associated with a modest increase in headache, blurred vision, depression and confusion. However, the frequency of these events was low overall. Heffron\textsuperscript{118} found no difference in overall frequency of adverse events but noted that drowsiness, lethargy and vertigo were more commonly reported with temazepam 30 mg. There is some evidence that temazepam 30 mg is associated with daytime impairment on tests such as the Choice Reaction Time Test and Critical Flicker Fusion Test (Hindmarch, 1979).

In summary, the task force found that there was weak evidence of efficacy of temazepam in terms of therapeutic effects on sleep initiation, total sleep time, awakenings, sleep efficiency, and possibly WASO with limited or no consistent evidence of adverse events in excess of placebo. However, there was also limited evidence for daytime impairment with temazepam 30 mg. Over, benefits were deemed to outweigh harms for temazepam 15 mg.

**Patients’ Values and Preferences**
Based on its clinical judgement, the task force determined that a majority of patients would be likely to use both temazepam 15 mg and 30 mg compared to no treatment.
4.7 Doxepin for the Treatment of Primary Insomnia

4.7a We suggest that clinicians use doxepin as a treatment for sleep maintenance insomnia (versus no treatment) in adults. [WEAK FOR]

Remarks: This recommendation is based on doxepin 3 mg and 6 mg.

4.7b Summary
Four studies addressed the efficacy of doxepin 3 mg. Four studies also investigated the 6 mg dosage. The overall quality of evidence for both dosages was LOW due to potential publication bias and imprecision. The evidence suggests minimal improvement in SL but clinically significant improvements in WASO, TST and SE. The overall evidence was weakly in favor of doxepin’s efficacy in improving sleep maintenance.

Meta-analysis shows that PSG and patient-reported SL at 3 mg and PSG SL at 6 mg fell well below the clinical significance threshold. The TST data demonstrate that both PSG and subjective TST at 3 mg, as well as PSG TST at 6 mg, were well above significance thresholds, although subjective TST at 6 mg fell somewhat short of this. The PSG data for reduction of WASO exceeded the clinical significance threshold at both dosages, although patient diary data for WASO at the 6 mg dosage fell substantially below threshold, based on two studies. The standardized mean difference (SMD) in sleep quality for doxepin 3 mg suggests moderate improvement, while the SMD for the 6 mg dosage was in a mildly improved range. The objective SE for both dosages exceeded the clinical significance level, while objective NOA fell well short.

Meta-analysis of side effects included headache, diarrhea, somnolence and upper respiratory infection at 3 mg, and headache and somnolence at the 6 mg dose. Results suggest mild to moderate increase in headache risk at both dosages and mild increase in somnolence at 6mg. Given the demonstrated improvements in WASO, TST and SE, with limited adverse effects, the task force judged the benefits to outweigh the harms. It was determined by clinical judgement of the task force that the majority of well-informed patients would use doxepin over no treatment. This judgement is based on the evidence for clinically significant improvement in WASO, TST and SE.


4.7c Discussion
A total of five studies investigated the effects of doxepin at 3 mg and/or 6 mg. Krystal conducted a 12 week RCT of nightly doxepin 1 and 3 mg versus placebo in 240 elderly (>65 years) subjects with predominant sleep maintenance insomnia. Subjects were randomized to one of three treatment groups. Outcome variables included both PSG and sleep diary data. Krystal investigated doxepin 3 mg and 6 mg in a five week trial which included 221 adults with sleep onset and maintenance insomnia who were randomized to one of the two doxepin doses or placebo. PSG data and sleep diaries were included. Roth employed a crossover design with randomized assignment to one of four treatment sequences which consisted of two nights of doxepin 1 mg, doxepin 3 mg, doxepin 6 mg and placebo, with intervening washout. PSG and sleep diary data were collected. The study included 67 adults who met both baseline PSG-defined sleep onset and maintenance criteria. Scharf employed the identical crossover design and dosages in 76 elderly insomnia subjects. Lankford reported data on a four week nightly trial of doxepin 6 mg or placebo in 254 elderly subjects with sleep onset and sleep maintenance insomnia. Outcome variables were patient-reported and clinician rated.
Hajak also conducted a RCT of doxepin, but the dosages (25-50 mg) were significantly higher than FDA-approved hypnotic dosages. For this reason, the study is not included in the discussion.

**Sleep latency**

Four studies reported PSG SL data for the 3 mg dosage. The mean difference from placebo (-2.30 min; CI: -6.22 to +1.62 min) was far below the defined significance threshold. Evidence quality was MODERATE due to potential publication bias. Likewise, patient-reported SL did not meet clinical significance (-9.35 min; CI: -21.89 to +3.19 min). Quality was LOW due to imprecision and potential publication bias. Three studies included adequate data for meta-analysis at the 6 mg dosage, showing a mean difference for objective SL of -5.29 min (CI: -1.34 to -9.25 min) with MODERATE quality of evidence due to publication bias. No sleep diary data were available for meta-analysis of SL at this dosage.

**Total sleep time**

Four investigations reported polysomnographic data for TST at 3 mg. The analysis reveals a clinically significant increase in TST at this dosage (+26.14 min; CI: +18.49 to +33.79 min). Quality was LOW due to imprecision and potential publication bias. Subjective reports for 3 mg are also well into the range of clinical significance (+43.57 min; CI: +5.16 to +81.98 min) with VERY LOW quality of evidence due to imprecision and potential publication bias. At the 6 mg dosage, PSG-determined TST, also met the clinical significance criterion (+32.27 min; CI: +24.24 to +40.30 min) with MODERATE quality of evidence due to potential publication bias. However, subjective TST at this dosage fell short of significance (+18.84 min; CI: -1.65 to +39.34 min) with LOW quality of evidence due to imprecision and potential publication bias.

**Wake after sleep onset**

WASO was considered a key outcome variable in all of the doxepin studies noted. The PSG data for 3 mg doxepin shows a clinically significant mean difference from placebo of -22.17 min (CI: -14.72 to -29.62 min), based on four trials. Quality of evidence was LOW due to imprecision and potential publication bias. Only one study reported subjective WASO, with a reduction of 20.0 minutes versus placebo. Quality of these data was LOW due to imprecision and potential publication bias. At 6 mg, PSG WASO showed a clinically significant reduction of 23.14 min (CI: -16.36 to -30.34 min) with LOW quality of evidence due to imprecision and potential publication bias. Patient diary results did not meet clinical significance (-14.39 min; CI: -3.93 to -24.86 min) with MODERATE quality of evidence due to potential publication bias.

**Quality of sleep**

Quality of sleep ratings for the 3 mg dosage suggest substantial improvement (SMD: +0.57; CI: +0.26 to 0.88 SMD) with LOW quality of evidence due to imprecision and potential publication bias, with more modest improvement at 6 mg (SMD: +0.28; CI: +0.06 to 0.49 SMD) with MODERATE quality of evidence due to potential publication bias.

**Sleep efficiency**

Polysomnographic SE was reported in three studies for the 3 mg dosage. Evidence quality was LOW due to imprecision and potential publication bias. The improvement in SE was clinically significant at +6.78% (CI: +4.50 to 9.07%). SE at the 6 mg dose, based on two investigations was also significantly improved (+7.06%; CI: +5.12 to 9.01%) with MODERATE quality of evidence due to potential publication bias.
**Number of awakenings**
PSG-determined NOA was mildly increased (+0.53 awakenings; CI: -0.37 to +1.42 awakenings) for 3 mg\textsuperscript{121, 123, 124} and the 6 mg dose (+0.44 awakenings; CI: -0.57 to +1.44 awakenings), with MODERATE quality for both, due to potential publication bias.

**Overall Quality of Evidence**
The quality of evidence in the meta-analytic data for the majority of variables was MODERATE to LOW due to industry sponsorship and, in some cases, imprecision (due to relatively large confidence intervals for numerous variables). Quality was further downgraded to VERY LOW for subjective TST at 6 mg; as a result of the above factors plus heterogeneity of data. As a result, the overall QOE for the doxepin data is considered VERY LOW.

**Harms**
Meta-analysis for headache was available for both the 3 mg\textsuperscript{121-123} and 6 mg\textsuperscript{122, 123, 125} dosages. This revealed a mild to moderate increased risk at both levels (0.06 and 0.05 risk difference versus placebo, respectively). Somnolence showed no significant increase vs. placebo (+0.01 risk difference) at the 3 mg level\textsuperscript{121-123} and a very mild increased risk at 6 mg (+0.04 risk difference)\textsuperscript{122, 123, 125}. Data was also available for meta-analysis of risk for diarrhea and upper respiratory infection. Neither showed significantly greater risk than placebo. With respect to next-day residual effects, no difference was observed between doxepin 3 mg or 6 mg and placebo on DSST, Symbol Copying Test (SCT) or visual analogue scales for morning sleepiness\textsuperscript{121-124}.

In summary, the task force found that there was weak evidence for efficacy in the treatment of sleep maintenance insomnia, with minimal evidence of adverse events in excess of placebo. Therefore, benefits were deemed to be greater than harms.

**Patients’ Values and Preferences**
Based on its clinical judgement, the task force determined that in light of the data supporting efficacy for reducing WASO, and improving TST, SE and sleep quality, a majority of patients would be likely to use doxepin compared to no treatment.

### 4.8 Trazodone for the Treatment of Primary Insomnia

**4.8a** We suggest that clinicians not use trazodone as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

**Remarks:** This recommendation is based on trazodone 50 mg.

See **Supplemental Materials**, Table S14.

**4.8b Summary**
A single study\textsuperscript{81} of trazodone 50 mg met inclusion criteria; therefore, no meta-analysis is available. The overall quality of evidence for this study was MODERATE due to potential publication bias. The patient-reported data from this study demonstrate a modest reduction in SL which falls substantially below the threshold for clinical significance. Likewise, the moderate increase in TST is not clinically significant and the reduction in WASO is minimal. Quality of sleep is insignificantly improved and reduction in NOAs falls just below clinical significance. In summary, none of the sleep outcome variables improved to a clinically significant degree.
No meta-analysis of harms was possible. Given the absence of demonstrated efficacy on numerous critical outcome variables, coupled with limited evidence regarding harms, the task force judged the harms to potentially outweigh the benefits. Based on its clinical judgement, the task force determined that, despite the absence of significant efficacy for trazodone 50 mg and the paucity of information regarding harms, the majority of patients would be likely to use trazodone 50 mg compared to no treatment.

4.8c Discussion
Walsh\textsuperscript{81} investigated the efficacy of trazodone 50 mg versus zolpidem 10 mg and placebo. The final sample for the trazodone and placebo groups included 187 adults with sleep onset insomnia. Subjects were administered either trazodone or placebo in double-blind fashion for 14 consecutive nights. All data were patient-reported.

Sleep latency
Subjective SL was reduced by 10.2 min (CI: -8.95 to -11.44 min). This falls short of the clinical significance threshold. The quality of evidence was MODERATE due to potential publication bias.

Total sleep time
Sleep diary TST was increased by a clinically insignificant 21.8 min (CI: +20.10 to +23.49 min). The quality of evidence was MODERATE due to potential publication bias.

Wake after sleep onset
Sleep diary WASO was reduced by 7.7 min (CI: -8.89 to -6.5 min), falling well below the threshold. The quality of evidence was MODERATE due to potential publication bias.

Quality of sleep
On a 4-point scale (1="excellent", 4="poor") sleep quality was insignificantly improved (-0.13 points; CI: -0.11 to -0.14 points). The quality of evidence was MODERATE due to potential publication bias.

Number of awakenings
This outcome was reduced by 0.4 (CI: -0.37 to -0.42 awakenings) compared to placebo, less than the 0.5 subjective awakening threshold. The quality of evidence was MODERATE due to potential publication bias.

Overall Quality of evidence
The overall quality of evidence for this study was MODERATE.

Harms
There was no meta-analysis of harms. In the Walsh\textsuperscript{81} paper, the trazodone group experienced significantly more side effects than the placebo group. Chief among these were headache (trazodone 30%; placebo 19%) and somnolence (trazodone 23%; placebo 8%). In all, 75% of trazodone subjects reported some adverse event(s), compared to 65.4% of subjects who received placebo.

Patients’ Values and Preferences
Based on its clinical judgement, the task force determined that, despite the absence of significant efficacy for trazodone 50 mg and the paucity of information regarding harms, the majority of patients would be likely to use trazodone 50 mg compared to no treatment. This is based on the perception of trazodone as a “safer” sleep promoting agent by many physicians and the resulting recommendations and prescribing practices of those physicians.
Anticonvulsants

4.9 Tiagabine for the Treatment of Primary Insomnia

4.9a We suggest that clinicians not use tiagabine as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

Remarks: This recommendation is based on tiagabine 4 mg.

4.9b Summary

Three studies addressed the efficacy of tiagabine 4 mg. The overall quality of evidence was VERY LOW due to potential publication bias, heterogeneity, and imprecision. Meta-analyses were conducted for SL (PSG and subjective), TST (PSG and subjective), WASO (PSG and subjective), sleep quality, SE (PSG), and NOA (PSG and subjective). These analyses revealed that both objective and subjective measures of sleep latency fell well below the threshold for clinical significance. Measures of TST showed minimal change (PSG) and mild to moderate reduction (sleep diary). WASO data demonstrated no clinically significant change on either metric. Meta-analysis of standardized mean difference for sleep quality suggested improvement which fell just below the clinical significance threshold. Neither objective nor subjective NOAs were reduced by clinically significant levels, while PSG SE was minimally reduced.

Meta-analysis of adverse effects showed no difference between tiagabine and placebo on headache or nausea. Given the absence of demonstrated efficacy on numerous critical outcome variables (with slight trending toward mild worsening on some outcomes), coupled with limited evidence regarding harms, the task force judged the harms to potentially outweigh the benefits.

It was determined by clinical judgement of the task force that the majority of well-informed patients would not use tiagabine over no treatment. This judgement is based on the lack of evidence for efficacy and the limited systematic information regarding adverse effects.

See Supplemental Materials; Figures S55-S64 and S84-S85, and Tables S15-S17.

4.9c Discussion

Three studies were included in the meta-analyses of tiagabine. Roth studied 207 elderly primary insomnia patients (65-85 years) with difficulty initiating and maintaining sleep, who received tiagabine 2, 4, 6, or 8 mg or placebo on two consecutive study nights with PSG recordings in a parallel group design. Walsh similarly evaluated 232 adults with chronic sleep-onset and maintenance insomnia. Tiagabine 4, 6, 8, or 10 mg or placebo was administered on two consecutive nights with polysomnography. Walsh conducted a crossover study of 58 adults (age 35-64) with chronic sleep initiation and maintenance problems. Subjects received 4, 8, 12, and 16 mg and placebo for two consecutive nights of sleep recording. Medication-free washout periods between doses ranged from 5-12 nights.

Sleep latency

The meta-analysis for SL included three studies. PSG SL data showed a minimal, clinically insignificant increase in SL (+3.65 min; CI: -8.00 to +15.31 min) with VERY LOW quality of evidence due to heterogeneity, imprecision and potential publication bias. The subjective data showed a moderate increase in SL (+13.31 min; CI: +7.54 to 19.37 min). Quality of evidence was MODERATE due to potential publication bias.
**Total sleep time**
Objective data for TST\textsuperscript{127-129} demonstrated a minimal *reduction* in TST (-1.21 min; CI: -7.44 to +5.02 min) with LOW quality evidence due to heterogeneity and potential publication bias. Patient-reported TST\textsuperscript{127,129} was *reduced* by 19.47 minutes (CI: -24.72 to -14.23 min) with MODERATE quality of evidence due to potential publication bias. Neither subjective nor objective findings met clinical significance.

**Wake after sleep onset**
The PSG WASO analysis\textsuperscript{127-129} revealed essentially no difference from placebo (-0.56 min; CI: -6.77 to +5.65 min). Quality of evidence was LOW due to heterogeneity and potential publication bias. Sleep diary data\textsuperscript{127,129} indicated a small, clinically insignificant *increase* (+4.29 min; CI: -0.22 lower to +8.79 min) with MODERATE quality of evidence due to potential publication bias.

**Quality of sleep**
The meta-analysis for QOS\textsuperscript{127,129} resulted in a standardized mean difference of +0.48 (CI: -0.5 to +1.46 SMD), which falls just below the level of clinical significance. Quality of evidence was VERY LOW due to heterogeneity, imprecision and potential publication bias.

**Sleep efficiency**
The objective sleep efficiency was *reduced* insignificantly (-0.53%; CI: -0.02 to -1.05%). Quality of evidence was MODERATE due to potential publication bias.

**Number of awakenings**
The PSG NOAs were mildly *increased* (+0.5 awakenings; CI: -1.29 to +2.29 awakenings). The subjective NOA was minimally reduced at -0.21 awakenings (CI: -0.9 to +0.48 awakenings), falling below the threshold for clinical significance. Level of evidence was LOW for both measures due to imprecision and potential publication bias.

**Overall Quality of evidence**
The overall quality of evidence for the meta-analytic data was VERY LOW due to significant heterogeneity, imprecision and industry sponsorship for some critical outcomes.

**Harms**
Meta-analysis was possible for two adverse effects (headache and nausea). Neither showed any significant difference from placebo. None of the three studies found a significant difference from placebo on morning-after DSST or visual analogue scales for sleepiness/alertness at the 4 mg dose.

**Patients’ Values and Preferences**
Based on its clinical judgement, the task force determined that in light of the absence of significant efficacy at this dose and the paucity of information regarding harms, the majority of patients would not be likely to use tiagabine 4 mg compared to no treatment.
4.10 Diphenhydramine for the Treatment of Primary Insomnia

4.10a We suggest that clinicians not use diphenhydramine as a treatment for sleep-onset and sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

Remarks: This recommendation is based on diphenhydramine 50 mg.

4.10b Summary
Two RCTs evaluated diphenhydramine 50 mg for the treatment of chronic primary insomnia. The overall quality of evidence was downgraded to LOW due to imprecision and risk of publication bias. The overall evidence for diphenhydramine 50 mg was weakly against its effectiveness for improving sleep initiation and TST. The mean reduction in patient-reported sleep latency versus placebo fell well below the level of clinically significant improvement. The same studies found a small increase in TST which also fell substantially below the threshold for significance. The single paper which included PSG-determined SL and TST showed outcomes which also fell mildly and moderately below clinical significance thresholds, respectively. None of the other objective or patient-reported outcome variables reached clinical significance. In addition, one paper meeting inclusion criteria but not including suitable data for meta-analysis evaluated diphenhydramine 50 mg in mild to moderate insomnia patients.

No meta-analysis was possible for side effects. Since no systematic data addressing harms is available, it is difficult to make a clear determination regarding benefits versus harms. However, in light of the absence of clear benefits, the task force judged the benefits and harms to be approximately equal. It was determined by clinical judgement of the task force that the majority of well-informed patients would not use diphenhydramine 50 mg over no treatment. This judgement is based on the absence of evidence for clinically significant improvement.

See Supplemental Materials; Figures S65-S66 and Table S18.

4.10c Discussion
Two studies of diphenhydramine 50 mg included adequate data for meta-analysis. Glass studied 25 elderly subjects (mean age = 73.9 years) with insomnia. Enrollees received diphenhydramine, temazepam 15 mg and placebo in a cross-over design with two weeks of nightly use for each intervention, followed by washout. Primary outcomes measures were sleep variables recorded in patient diaries. Morin compared diphenhydramine (14 nights, followed by 14 nights of placebo) to a valerian-hops preparation (28 nights) and placebo (28 nights) in a total population of 184 adults with occasional insomnia (2-4 nights/week with SL >30 min or WASO >30 min). Patients were randomized to one of the three intervention groups and PSG and patient-reported data were included. A third study, not included in meta-analysis, assessed mild to moderate insomnia patients in family practice settings, every subject receiving diphenhydramine 50 mg and placebo for one week each in crossover fashion, without intervening washout. Outcome assessment was based on patient-completed sleep questionnaires.

Sleep latency
The single study employing PSG found a 7.89 min reduction in SL (CI: -17.40 to +1.62 min). This fell far below the significance threshold. Quality of evidence was LOW due to imprecision and potential publication bias. Two studies met requirements for meta-analysis of subjective SL. This revealed a mean difference from placebo of -2.47 min (CI: -8.17 to +3.23 min). The Rickels study found statistically significant improvement in SL with diphenhydramine using a 0-4 patient-rating scale, but no specific quantitative data regarding actual SL times is included.
Total sleep time
Morin\textsuperscript{130} reported a PSG TST increase of 12.37 min (CI: -13.38 to +38.12 min). This falls well below the significance threshold of 20m. Quality of evidence was LOW due to imprecision and potential publication bias. Meta-analysis of the two studies demonstrated a 17.86 min increase (CI: -3.79 to + 39.51 min) in subjective TST versus placebo. The Rickels study\textsuperscript{131} found “statistically significant improvement” in patient-reported TST but, as noted above, it is unclear to what extent this represents clinically significant improvement.

Wake after sleep onset
No data pertaining to wake after sleep onset was available.

Quality of sleep
Glass\textsuperscript{113} found minimal difference in sleep quality between diphenhydramine and placebo (mean difference of +0.1 SD; CI: -0.45 to +0.65 SD). Quality of evidence was downgraded to MODERATE due to potential publication bias. Rickels\textsuperscript{131} reported statistically significant improvement.

Sleep efficiency
The objective sleep efficiency data from the single study reporting PSG analysis\textsuperscript{130} found no clinically significant improvement (+2.59%; CI: -3.25 to +8.43%). In this same study, subjective SE also fell below the threshold (+4.61%; CI: +1.33 to +7.88%).

Number of awakenings
Subjective number of awakenings (-0.3 awakenings; CI: -1.03 to +0.43 awakenings) was not clinically significant.

Overall Quality of Evidence
The overall quality of evidence in the meta-analytic data from these studies was downgraded to LOW for imprecision, due to a relatively large confidence interval, which cross the clinical significance thresholds for subjective TST, a critical outcome. These studies were industry sponsored, resulting in further downgrading of evidence due to potential publication bias. The quality of evidence for individual critical outcomes ranged from moderate to low, therefore the overall quality of evidence was LOW.

Harms
No meta-analysis of adverse effects was possible. Neither Morin\textsuperscript{130} nor Glass\textsuperscript{113} found significant differences between diphenhydramine and placebo in adverse events. Rickels\textsuperscript{131} reported somewhat higher rates of drowsiness, dizziness and grogginess with diphenhydramine but no statistical analysis was conducted.

Morin\textsuperscript{130} found no substantial rebound effects following discontinuation of diphenhydramine. Glass\textsuperscript{113} noted minimal differences between diphenhydramine and placebo in the number of subjects experiencing rebound for at least one sleep outcome variable. Glass\textsuperscript{113} found no difference in morning-after DSST or Manual Tracking Task (MTT) between interventions.

In summary, the task force found that there was weak evidence for absence of efficacy in the treatment of sleep initiation insomnia, with minimal evidence of adverse events in excess of placebo. Therefore, benefits were deemed approximately equal to harms.
Patients' Values and Preferences
Based on its clinical judgement, the task force determined that in light of the paucity of data supporting efficacy for sleep initiation and maintenance, a majority of patients would not be likely to use diphenhydramine compared to no treatment.

4.11 Melatonin for the Treatment of Primary Insomnia

4.11a We suggest that clinicians not use melatonin as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

Remarks: This recommendation is based on melatonin 2 mg.

4.11b Summary

Three studies addressed the efficacy of melatonin 2 mg.132-134 These investigations included only older adults (>55 years). The overall quality of evidence was VERY LOW due to potential publication bias, heterogeneity, and imprecision. Meta-analysis was only achievable for sleep quality. This indicated a standardized mean difference of +0.21 (CI: -0.36 to +0.77 SMD), which was not clinically significant. The minimal overall evidence available was weakly against melatonin's efficacy in improving sleep onset, maintenance, or quality.

No adequate data for meta-analysis of adverse effects was available. Given the lack of evidence for efficacy in treating insomnia, and the unavailability of systematic data on side effects, the task force judged the benefits to be approximately equal to harms. It was determined by the task force that the majority of well-informed patients would use melatonin over no treatment. This is based on its availability and the widespread perception of melatonin as a benign sleep aid.

See Supplemental Materials; Figures S67 and Table S19.

4.11c Discussion

Three studies included adequate data for melatonin meta-analysis.132-134 Lemoine132 studied 170 older adults (age >55 years) with primary insomnia. Subjects received either prolonged release melatonin (PRM) 2 mg or placebo nightly for 3 weeks. Outcome data was patient-reported. Luthringer133 similarly studied 40 older adults (age >55 years) who received PRM 2 mg or placebo for 3 weeks. Outcomes included both PSG and subjective data. Finally, Wade134 evaluated 354 patients of the same age group with PRM 2 mg or placebo nightly for 3 weeks. Outcome data was patient-reported.

In addition, seven trials which met inclusion criteria but did not include adequate data for meta-analysis were identified.135-141 These investigations employed various dosages and combinations with other agents, rendering meaningful comparisons to the 2 mg RCTs impossible. Pertinent features of these studies are included within each outcome section.

Haimov137 conducted a randomized cross-over study of elderly adults with insomnia consisting of one week on each of three interventions (2 mg sustained-release melatonin, 2 mg fast-release melatonin or placebo) with intervening washout, followed by a 2 month extension of 1 mg slow-release melatonin. Data were derived from actigraphy. Zhdanova141 evaluated three dosages of melatonin (0.1, 0.3, and 3 mg) versus placebo in a randomized crossover study of 30 elderly (>50 years) adults (15 normal sleepers and 15 insomnia subjects with reduced SE). Subjects received each dosage or placebo for one week with
intervening washout. Wade\textsuperscript{140} administered prolonged-release melatonin 2 mg or placebo to adults with primary insomnia for 3 weeks, following which the melatonin group continued for 26 weeks, while the placebo group was re-randomized to melatonin or placebo (1:1). Sleep outcome variables (from sleep diary) were analyzed according by age group as well as by melatonin deficiency status. Baskett\textsuperscript{135} conducted a randomized controlled crossover study of healthy elderly with sleep maintenance problems. Subjects received 5 mg melatonin or placebo for four weeks with intervening washout.

**Sleep latency**
Meta-analysis was not possible for sleep latency. Luthringer\textsuperscript{133} reported a PSG SL reduction of 8.9 min (CI: -2.35 to -15.45 min), which falls below clinical significance (prolonged release 2 mg). The quality of evidence was LOW due to imprecision and potential publication bias.

In the Haimov\textsuperscript{137} investigation, fast-release melatonin produced significantly shorter SL than placebo at one week. At 2 months, sustained release 1 mg resulted in significantly shorter SL than placebo. Zhadanova\textsuperscript{141} reported no significant improvement in PSG SL at any dosage.

Wade\textsuperscript{140} found that the melatonin deficient group (including all ages) showed no improvement with melatonin versus placebo on SL at three weeks. However, the elder group (65-80 years) showed significant reduction of SL with melatonin, regardless of melatonin deficiency status (SL: -19.1 min; placebo -1.7 min). This improvement held at 19 weeks for the elder group (melatonin: -25.9 min; placebo: -8.3 min). Wade\textsuperscript{139} subsequently re-analyzed these data and reported that the significant improvement in SL held when the age range for the “elderly” group was expanded to 55-80 years, but not lower. Baskett\textsuperscript{135} found no improvement in SL (sleep diary) with melatonin 5 mg.

**Total sleep time**
There were inadequate data for meta-analysis of TST. Luthringer\textsuperscript{133} found an insignificant change (+2.2 min; CI: -19.13 to +23.53 min) in objective TST with melatonin 2 mg. The quality of evidence was VERY LOW due to significant imprecision of the data, and potential publication bias.

Zhdanova\textsuperscript{141} observed no increase in objective TST at any dosage. Wade\textsuperscript{140} reported no improvement in patient-reported TST in the low melatonin excretor population (regardless of age) at 3 weeks but observed a small improvement (estimated difference: +13.1 min) at 29 weeks. Analysis of the elderly population revealed no significant improvement in TST at any point. Baskett\textsuperscript{135} reported no improvement at the 5 mg dose as measured by sleep diary.

**Wake after sleep onset**
No meta-analysis was possible for WASO. Luthringer\textsuperscript{133} found a small increase in WASO (+8.5 min; CI: -11.75 to +28.75 min) in the prolonged release melatonin 2 mg group. The quality of evidence was VERY LOW due to significant imprecision of the data, and potential publication bias.

**Quality of sleep**
The meta-analysis of QOS demonstrated a small improvement in quality of sleep (+0.21 SMD; CI: -0.36 to +0.77 SMD), which falls well below the threshold for clinical significance. The quality of evidence was VERY LOW due to heterogeneity, imprecision and potential publication bias.
Baskett\textsuperscript{135} found no improvement in quality of sleep with 5 mg melatonin. Wade\textsuperscript{140} reported no improvement with prolonged-release melatonin at 3 weeks and 29 weeks in the low excretor and elderly groups.

**Sleep efficiency**
There were not adequate data for meta-analysis of melatonin SE.

Haimov\textsuperscript{137} reported small to moderate increases in actigraphic SE versus placebo (placebo: 77.4%; fast-release 2 mg/1 week: 78.8%; sustained release 2 mg/1 week: 80.4%; sustained release 1 mg/2 months: 84.3%). Both of the sustained release dosages and durations were statistically significantly different from placebo. Zhdanova\textsuperscript{141} also reported significant improvement in PSG SE versus placebo in the multiple dose crossover study (placebo: 78%; melatonin 0.1 mg: 84%; 0.3 mg 88%; 3 mg: 84%). Baskett\textsuperscript{135} found no difference between placebo and melatonin 5 mg for subjective SE.

**Number of awakenings**
Insufficient data precluded meta-analysis of NOA. Luthringer\textsuperscript{133} found an increased (+1.4 awakenings; CI: -4.59 to +7.39 awakenings) NOA (PSG) with melatonin. The quality of evidence was VERY LOW due to significant imprecision of the data, and potential publication bias.

Zhdanova\textsuperscript{141} and Baskett\textsuperscript{135} reported no difference in NOA between melatonin and placebo by PSG or patient diary, respectively.

**Overall Quality of Evidence**
The overall quality of evidence in the single outcome meta-analytic data from these studies was downgraded to very low due to heterogeneity, imprecision, and industry sponsorship, resulting in potential publication bias.

**Harms**
Meta-analysis for adverse events was not possible. Of the included investigations, none reported clinically significant differences in adverse events between melatonin and placebo for any dosage or duration.\textsuperscript{132-134} With one possible exception, no rebound or withdrawal effects were reported.\textsuperscript{132, 133, 140} Haimov\textsuperscript{137} found marginally significant difference in SE between the active phase for two month, 1 mg sustained-release melatonin and the withdrawal period.

In summary, the task force found that there was weak evidence against clinically significant efficacy in the treatment of sleep onset insomnia, with little systematic evidence regarding harms. However, mixed evidence suggests possible improvement in SL in an elderly population. Therefore, benefits were deemed to be approximately equal to harms.

**Patients’ Values and Preferences**
Based on clinical judgement, the task force determined that despite the paucity of meta-analytic data, equivocal data regarding efficacy for sleep-onset insomnia, and absence of data regarding sleep maintenance, a majority of informed patients would be likely to use melatonin compared to no treatment. As previously noted, this is based on its availability and the widespread perception of melatonin as a benign sleep aid.
4.12 L-tryptophan for the Treatment of Primary Insomnia

4.12a We suggest that clinicians not use tryptophan as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

Remarks: This recommendation is based on tryptophan 250 mg.

4.12b Summary
Only one study\textsuperscript{142} contained data sufficient for meta-analysis which was, therefore, not possible. The quality of evidence for this study was MODERATE. This study, consisting of patient-reported data, shows a modest decline in TST, which was not clinically significant. WASO was decreased slightly, while sleep quality was mildly increased; neither were of clinical significance. Sleep efficiency was insignificantly decreased.

No meta-analysis of harms was possible. Given the absence of demonstrated efficacy on numerous critical outcome variables, coupled with limited evidence regarding harms, the task force judged the harms to potentially outweigh the benefits. Based on its clinical judgement, the task force determined that, despite the absence of significant efficacy for tryptophan 250 mg and the absence of information regarding harms, the majority of patients would be likely to use tryptophan 250 mg compared to no treatment.

See Supplemental Materials; Table S20.

4.12c Discussion
Hudson\textsuperscript{142} investigated the effects of food source tryptophan (250 mg), pharmacological tryptophan 250 mg, both with carbohydrate, versus carbohydrate alone. Subjects (N=31) received one of the three interventions for one week. Outcome data consisted of sleep diaries.

Two additional papers met inclusion criteria, but used much higher dosages. Hartmann\textsuperscript{143} compared tryptophan 1 g to secobarbital, flurazepam, and placebo in a one week trial. Tryptophan and placebo groups included 52 subjects with chronic insomnia. Data were patient-reported. Spinweber\textsuperscript{144} studied 20 young men with sleep onset insomnia. Following placebo run-in, ten subjects received tryptophan 3 g and ten received placebo for six nights, with PSG recordings nightly.

Sleep latency
The Hudson\textsuperscript{142} study did not report sleep latency data. Hartmann\textsuperscript{143} found no difference between tryptophan and placebo during active treatment. Spinweber\textsuperscript{144} noted improvement in sleep latency only on nights 4-6 of administration (11.2 min lower than placebo for this period).

Total sleep time
Hudson\textsuperscript{142} reported a moderate reduction in TST (-20 min; CI: -31.29 to -8.7 min). The quality of evidence was MODERATE due to imprecision. Other investigations did not report TST data.

Wake after sleep onset
Hudson\textsuperscript{142} noted a small reduction in WASO (-9.7 min; CI -15.21 to -4.18 min), that did not meet clinical significance. The quality of evidence was HIGH.
Quality of sleep
On a 3-point scale (1=low, 3=high) sleep quality was mildly increased (+0.3 points: CI +0.22 to +0.37 points). The quality of evidence was HIGH. Hartmann found no significant difference between tryptophan and placebo on a measure of "How well I slept."

Sleep efficiency
Sleep efficiency was not reported by any study.

Number of awakenings
NOA was not reported by any study.

Overall Quality of evidence
The overall quality of evidence for this study was MODERATE.

Harms
There was no meta-analysis of harms. None of the papers reported systematic information regarding adverse effects associated with tryptophan.

Patients’ Values and Preferences
Based on clinical judgement, the task force determined that, despite the absence of significant efficacy for tryptophan 250 mg and the absence of information regarding harms, the majority of patients would be likely to use tryptophan 250 mg compared to no treatment.

4.13 Valerian for the Treatment of Primary Insomnia

4.13a We suggest that clinicians not use valerian as a treatment for sleep onset or sleep maintenance insomnia (versus no treatment) in adults. [WEAK AGAINST]

Remarks: This recommendation is based on variable dosages of valerian and valerian-hops combination.

4.13b Summary
Morin evaluated a combination of valerian (374 mg native extract) and hops (83.8 native extract). The overall quality of evidence for these data was VERY LOW due to imprecision and potential publication bias. PSG sleep latency was reduced to a degree which fell just below the clinical significance threshold. Other measures, including subjective SL, as well as PSG and patient-reported TST and SE were improved to clinically insignificant degrees.

No meta-analysis of harms was possible. Given the absence of demonstrated efficacy on critical outcome variables (with the possible exception of marginally improved PSG SL), coupled with limited evidence regarding harms, the task force judged the harms to be roughly equal to the benefits. Based on its clinical judgement, the task force determined that, given the lack of efficacy for valerian (with the possible exception of small improvements in SL) and the limited information regarding harms, the majority of patients would not be likely to use valerian compared to no treatment.

See Supplemental Materials; Table S21.
4.13c Discussion
Morin\textsuperscript{130} investigated the effects of a valerian-hops combination in dosages noted above. This combination was compared to diphenhydramine and placebo. Subjects with mild difficulty initiating or maintaining sleep were randomized to one of the three interventions (valerian-hops N=59; placebo N=65) with nightly administration for 28 days. A subset (valerian N=22; placebo N=26) underwent PSG at baseline and end of weeks one and two.

One additional paper\textsuperscript{145} met inclusion criteria, but employed a higher dosage. Oxman conducted a randomized trial involving 405 adults of all ages with insomnia. Subjects were randomized to two-week, nightly administration of valerian (3600 mg) or placebo. Outcomes were patient-reported and captured as a ranges, therefore the data was not usable for meta-analysis.

**Sleep latency**
Morin\textsuperscript{130} found a reduction in PSG SL of 9.29 min (CI: -0.27 to -18.3 min). This approached clinical significance. The quality of evidence was LOW due to imprecision and potential publication bias. The subjective SL, however, was minimally increased at +3.77 min (CI: -4.47 to +12.01 min), with MODERATE quality of evidence due to potential publication bias. Oxman\textsuperscript{145} found no statistically significant improvement in SL.

**Total sleep time**
In the Morin\textsuperscript{130} study, PSG TST was increased, although not to a clinically significant degree (+10.96 min; CI: -21.67 to +43.59 min) (VERY LOW quality of evidence). Patient-reported TST was minimally higher (+3.12 min; CI: -22.08 to +28.32) with MODERATE quality of evidence. Oxman\textsuperscript{145} found no significant improvement in subjective TST.

**Wake after sleep onset**
WASO data were not reported in any study.

**Quality of sleep**
Morin\textsuperscript{130} did not report quality of sleep data and Oxman\textsuperscript{145} found no statistically significant improvement.

**Sleep efficiency**
Minimal increases in objective (+0.96%; CI: -5.02 to +6.94%) and subjective (1.85%; CI: -1.9 to +5.6%) SE were noted by Morin\textsuperscript{130} Both outcomes were downgraded due to potential publication bias, while PSG data was downgraded further due to significant imprecision.

**Number of awakenings**
Oxman\textsuperscript{145} observed a statistically significant reduction in average change scores for NOA with valerian.

**Overall Quality of evidence**
Quality of evidence for all outcomes ranged from very low to moderate, therefore the overall quality of evidence was VERY LOW.
**Harms**
Morin\textsuperscript{130} observed no difference between valerian-hops and placebo with respect to frequency of adverse events. No serious adverse events were noted. Likewise, Oxman\textsuperscript{145} found no increase in adverse events at the higher valerian dose compared to placebo.

**Patients’ Values and Preferences**
Based on its clinical judgement, the task force determined that, given the lack of efficacy for valerian (with the possible exception of small improvements in SL) and the limited information regarding harms, the majority of patients would not be likely to use valerian compared to no treatment.

### 5.0 LITERATURE REVIEWS

The following section contains literature reviews of drugs for which making clinical practice recommendations was not possible.

#### 5.1 Suvorexant

**Summary**
Two studies\textsuperscript{146, 147} evaluated suvorexant for treatment of chronic primary insomnia at FDA-approved dosages. The format of data reporting in these studies (difference between active treatment and placebo change-from-baseline) did not allow comparable meta-analysis for any sleep outcome variables. Meta-analysis was not possible for adverse events. Therefore, no recommendation for suvorexant was possible. The Herring 2012 study assessed 10 mg and 20 mg dosages while the Herring 2016 trial evaluated a 20 mg dosage in adults <65 years and 15 mg in older adults, but pooled the two dosages for analysis. Higher dosages evaluated in these trials (i.e. those exceeding FDA-approved maximum dosage recommendations) were disregarded. The results discussed are derived from the final assessment periods of these studies (4 weeks in Herring, 2012 and 3 months in Herring, 2016).

Overall, the objective data (PSG) demonstrate statistically significant improvement in latency to persistent sleep (LPS) at the 15 mg/20 mg dosages. Reduction in LPS at 10 mg was not statistically significant. Subjective time to sleep onset (TSO) was statistically significant in only one study (Herring, 2016) at the pooled 15/20 mg dosage although improvement was well below the threshold for clinical significance. The investigation of 10 mg and 20 mg in the Herring 2012 study found neither a statistically nor clinically significant improvement in TSO. PSG TST was reported in only one study\textsuperscript{147}. Statistically significant improvement was found at both 10 mg and 20 mg dosages. Subjective TST data were mixed, with one study\textsuperscript{147} demonstrating no statistically significant improvement at 10 mg and 20 mg. The other investigation found statistically significant improvement at the 15/20 mg dose, although outcomes varied with respect to the magnitude of improvement across the two separate trials included in the study.

PSG WASO data, for the most part, showed overall statistically significant reduction at all dosages evaluated. Subjective WASO was reported only by Herring, 2016. Reductions were small and failed to meet statistical significance. PSG sleep efficiency results were reported by Herring, 2012. Statistically significant improvement was found for both 10 mg and 20 mg. PSG number of awakenings (NAW) was insignificantly reduced or increased across both studies. Finally, sleep quality ratings showed minimal change. Adverse events were assessed in both studies. Overall frequency of adverse events was not significantly increased versus placebo. There was no evidence of daytime residual nor withdrawal symptoms.
Discussion
Herring 2012\textsuperscript{147} evaluated adults 18-64 years of age with DSM-IV primary insomnia in a randomized placebo-controlled cross-over study which included two 4-week trial periods. 62 subjects received 10 mg suvorexant and 61 received 20 mg. Subjects underwent PSG at the end of week 4. Sleep diary data was also obtained. The primary endpoint was sleep efficiency while secondary endpoints included latency to persistent sleep and wake after sleep onset. Inclusion criteria were LPS >20 min and WASO >60 min.

Herring 2016\textsuperscript{146} conducted two randomized placebo-controlled parallel trials of 3 months each. Adults 18-64 years-old as well as adults >65 with primary insomnia were included. 254 and 239 patients were randomized to suvorexant 15/20 mg in the two trials, respectively. The dosages of interest for this analysis were 20 mg for younger adults and 15 mg for the older population. Data for the two dosages were pooled for analysis. Inclusion criteria were LPS >20 min and WASO >60 min. Sleep diary data was collected for all patients and a subset underwent PSG. Both studies reported data as difference between placebo and drug change-from-baseline.

Sleep latency
Herring 2012\textsuperscript{147} found an insignificant reduction of 2.3 min change-from-baseline for suvorexant 10 mg compared to placebo. At the 20 mg dosage, a statistically significant reduction versus placebo of 22.3 min was reported. LPS in the two trials of Herring 2016 showed reductions of 8.1 and 0.3 min in the two trials, respectively. Neither of these reach the clinical significance threshold. Subjective TSO in the Herring 2012 study was insignificantly reduced at both dosages (-3.0 m and -4.3 m at 10 mg and 20 mg, respectively).

Herring 2016 reported statistically significant reduction in TSO at the pooled 15/20 mg dosages (-5.2 min and -7.6 min for the two trials).\textsuperscript{146}

Total sleep time
PSG TST was reported only in the Herring 2012 investigation. At both 10 mg and 20 mg, statistically significant improvement was seen versus placebo (+22.3 min and +49.9 min, respectively). Subjective TSO was minimally and insignificantly improved for both 10/20 mg in the Herring 2012 investigation (-3.0 min and -4.3 min).\textsuperscript{147} Statistically significant reduction of TSO was reported for both trials of the 15/20 mg dosages by Herring 2016 (trial 1: -5.2 min; trial 2: -7.6 min).\textsuperscript{146}

Wake after sleep onset
Both studies reported PSG WASO. Herring 2012 found statistically significant reduction of WASO at both 10 mg and 20 mg (-21.4 and -28.1 min, respectively).\textsuperscript{147} Herring 2016 reported reductions of -16.6 min and -31.1 min in the two trials, respectively.\textsuperscript{146} Reductions of subjective WASO in the two trials of 15/20 mg suvorexant in the Herring 2016 study reported mixed results where the second trial found statistically significant improvement (-7.7 min) but the first trial did not (-2.4 min).\textsuperscript{146}

Quality of sleep
Herring 2012 and 2016 reported mixed results for sleep quality. Herring 2012 reported small reductions in quality of sleep that were not statistically significant (-3.4 and -2.8 on a 100-point scale)\textsuperscript{147}, while Herring 2016 reported improvements in quality of sleep that met statistical significance in the second trial but not the first (0.1 points on a 4-point scale, for both trials).\textsuperscript{146}

Sleep efficiency
Herring 2012 found PSG SE improvement of +4.7% (10 mg) and +10.4% (20 mg). Herring 2016 did not report on sleep efficiency.\textsuperscript{147}
Number of awakenings

Number of awakenings showed no significant reduction in either study.

Adverse effects

Neither study found a significant increase in one or more adverse events versus placebo for suvorexant in the 10-20 mg range. Rates of serious adverse events were negligible and not significantly different between suvorexant and placebo. Frequency of daytime somnolence was somewhat increased in the 15/20 mg range (Herring, 2012: placebo 0.4%, 20 mg 4.9%; Herring, 2016 trial 1: placebo 3.4%; 15/20 mg = 5.1%; [trial1]; Herring, 2016 trial 2: placebo = 3.1%; 15/20 mg = 8.4%).[146, 147] The degree of somnolence was reported to be typically mild to moderate and did not often result in discontinuation. Frequency of somnolence was noted to increase in dose-dependent fashion at dosages exceeding FDA-recommended levels.

Assessments of withdrawal symptoms and daytime performance decrements did not reveal significant problems in either domain. There was no evidence of significant emergence of narcolepsy symptoms.

5.2 Estazolam

Summary

Three studies evaluated the efficacy of estazolam.[148-150] The three studies of the efficacy of estazolam utilized similar patient sleep questionnaires but none of the data are suitable for meta-analysis. Likewise, it is not possible to evaluate these data with respect to the established clinical significance thresholds. Therefore, no recommendations regarding efficacy of estazolam are possible. The data, such as it is, suggests statistically significant subjective improvement versus placebo at the 2 mg dosage for all parameters assessed.

In light of the absence of evidence for clinically significant improvement, the task force judged the potential harms to outweigh the benefits.

Discussion

Cohn[148] compared estazolam 1 mg and 2 mg to flurazepam and placebo in approximately 100 adults with chronic sleep onset and maintenance insomnia in a parallel group design. Subjects were randomized to receive drug or placebo for seven consecutive nights. Outcomes were measured by sleep questionnaires (interval ratings and Likert scales). Dominguez[149] evaluated a similar population of 45 adults with estazolam 2 mg, flurazepam or placebo for 7 nights. Sleep variables were assessed by patient questionnaire. Scharf[150] studied 243 outpatients with complaints of sleep initiation or maintenance difficulty. Subjects were randomly assigned to one of three parallel groups – estazolam 2 mg, flurazepam 30 mg or placebo. Treatment condition was administered for 7 nights. Subjects rated sleep latency, TST, QOS and NOA on numerical interval questionnaires.

Three studies found statistically significant improvement in SL on patient ratings with estazolam 2 mg. The one study which included estazolam 1 mg reported no significant improvement on SL. All three studies reported significant improvement versus placebo in sleep duration at 2 mg. The 1 mg dosage also produced significant improvement in sleep duration. Sleep quality was likewise improved at both dosages studied, as were NOA. No studies assessed WASO or SE.
5.3 Quazepam

Summary
Seven studies evaluated the efficacy of quazepam versus placebo in randomized, controlled trials.\textsuperscript{111, 151-156} One of these studies\textsuperscript{156} reported PSG findings while the remainder relied exclusively on subjective data derived from sleep questionnaires. Data analysis varies somewhat across these studies, rendering comparisons difficult. Only one investigation\textsuperscript{156} met requirements for meta-analysis, which was, therefore, not feasible. Overall, the studies are suggestive of efficacy in reducing time to onset of sleep, increasing TST, and reducing NOA, but, as stated, the methodologies employed are not comparable to the standard of data reporting required by GRADE and, therefore, it is difficult to make any specific recommendation. Quazepam and its metabolites have quite long half-lives, raising concerns regarding accumulation and daytime impairment. The data regarding daytime sleepiness from these studies suggests what may be a significantly higher percentage of patients with somnolence, particularly at the 30 mg dosage.

Discussion
Alden\textsuperscript{151} evaluated 57 insomnia subjects in a 5 night, parallel group design with quazepam 30 mg as the active drug. This study and all additional quazepam studies reported here (with the exception of Roth\textsuperscript{156}) utilized patient sleep questionnaire data consisting of numerical interval and other rating scales. Hernandez\textsuperscript{152} studied 36 insomnia outpatients with quazepam 15 mg and placebo in a similar five night design. Martinez\textsuperscript{153} assessed 41 older adults (> 65 years) with insomnia in a controlled trial with quazepam 15 mg or placebo administered over 5 consecutive nights. Mendels\textsuperscript{154} assessed the same dosage in 60 adult insomnia outpatients for five nights. O’Hair\textsuperscript{155} reported results of a five night trial in 60 subjects with quazepam 30 mg. Scharf\textsuperscript{111} studied quazepam 15 mg and triazolam 0.5 mg versus placebo over a five week period. During this time, subjects received active drug or placebo for nine consecutive nights, followed by 14 nights of every other night administration. Subjects were 65 insomnia outpatients. Finally, Roth\textsuperscript{156} evaluated quazepam 7.5 mg and 15 mg versus placebo in 30 older insomnia subjects (>60 years). PSG was conducted for two nights in the early phase of treatment (nights 1 and 2 of active treatment) and during the late phase (nights 6 and 7).

Sleep latency
Utilizing a cutoff of sleep latency <45 min to identify “responders,” Aden\textsuperscript{151} reported quazepam 30 mg to be statistically superior to placebo for % responders. O’Hair\textsuperscript{155} demonstrated quazepam 30 mg to be significantly better than placebo on an interval scale for sleep latency.

Hernandez\textsuperscript{152} found quazepam 15 mg significantly better than placebo on sleep latency interval scales. Likewise, Scharf\textsuperscript{111} reported significantly shorter latencies at this dosage on interval scales during active treatment nights in every-other-night administration although this was apparently not the case during the initial nightly administration. Using a 45 min sleep latency cutoff as described above\textsuperscript{151}, Martinez\textsuperscript{153} demonstrated a significantly higher percentage of responders to 15 mg in a geriatric population. Roth\textsuperscript{156} did not report significant differences between quazepam 7.5 mg or 15 mg and placebo on PSG SL.

Total sleep time
Utilizing a cutoff of sleep duration > 6hr to identify “responders,” Aden\textsuperscript{151} reported quazepam 30 mg to be statistically superior to placebo for % responders on 3/5 nights. O’Hair\textsuperscript{155} demonstrated quazepam 30 mg significantly better than placebo on an interval scale for TST.

Hernandez\textsuperscript{152} found quazepam 15 mg to be significantly superior to placebo on sleep duration interval scales. Likewise, Scharf\textsuperscript{111} reported significantly longer duration at this dosage on interval scales during active treatment nights in every-other-night administration although this was not the case during the initial nightly administration. Using a >6hr sleep duration cutoff as described above\textsuperscript{151}, Martinez\textsuperscript{153} demonstrated
a significantly higher percentage of responders to 15 mg in a geriatric population. Roth\textsuperscript{156} reported improvement in PSG TST during early (treatment nights 1 and 2) and late (nights 6 and 7) with quazepam 15 mg in a geriatric insomnia population. Significance was seen only during the late phase (nights 6 and 7) with quazepam 7.5 mg.

**Wake after sleep onset**
No studies reported placebo comparisons for WASO.

**Quality of sleep**
The majority of studies of “sleep quality” with quazepam utilized a composite index for sleep quality (including questions on nightmares and overall evaluation of the medication) which is not consistent with sleep quality measures used in other studies; these results are, therefore, not discussed. Scharf\textsuperscript{111} reported a single measure of sleep quality (“How would you describe your sleep”) with quazepam 15 mg to be significantly better than placebo on active treatment nights in both the nightly and every other night administration.

**Sleep efficiency**
No studies reported placebo comparison data for SE.

**Number of awakenings**
Employing a threshold for “response” of <2 awakenings, Aden\textsuperscript{151, 155} reported a significantly higher percentage of responders to quazepam 30 mg than placebo. O’Hair also found significantly fewer awakenings at this dosage compared to placebo using interval scales. At the 15 mg dosage, two studies\textsuperscript{152, 153} found a significantly greater number of “responders” (i.e. <2 awakenings) compared to placebo. No PSG data for NOA were reported.

**Adverse events**
Five studies reported specific data for daytime somnolence. Aden\textsuperscript{151} found an approximately four-fold higher rate of somnolence at 30 mg (quazepam 16/24; placebo = 4/26). At the same dose, O’Hair\textsuperscript{155} reported somnolence in 12/30 quazepam and 5/30 placebo subjects. At 15 mg, Martínez\textsuperscript{153} noted no difference in adverse events. Hernández\textsuperscript{152} reported somnolence in 9/30 quazepam subjects and 6/30 placebo subjects. Mendels\textsuperscript{154} found 7/30 quazepam subjects and 4/30 placebo subjects demonstrated daytime somnolence.

### 5.4 Flurazepam

**Summary**
Sixteen studies met general inclusion and exclusion criteria.\textsuperscript{101, 103, 104, 108, 112, 143, 148-150, 157-163} No studies contained data sufficient for meta-analysis. No meta-analysis of harms was possible. These studies were highly varied in design. Of these, three\textsuperscript{103, 104, 108} included no flurazepam/placebo comparison and were excluded from discussion. All of the studies included one or both of the standard flurazepam doses: 15 mg and 30 mg.

Studies of the efficacy of flurazepam are fraught with numerous methodological inconsistencies, not the least of which is that the instruments employed for subjective assessments of sleep outcomes are highly varied across these studies, casting doubt on the extent to which valid comparisons can be made across studies. Many incorporated interval scales with no reports of specific values. In light of these inconsistencies, and the related unavailability of meta-analyses, no recommendations regarding efficacy of flurazepam are possible. The data for sleep onset at both the 15 mg and 30 mg dosages are mixed. The majority of studies did report increases in TST with the 30 mg dosage but this trend is not evident at 15 mg.
Data for WASO are limited to two studies, one of which (the PSG study) showed improvement at 30 mg. Sleep quality reports uniformly indicated improvement at both dosages, while reports for NOA suggest reduction at the 30 mg dosage only.

**Discussion**

Cohn\textsuperscript{148} compared flurazepam 30 mg and placebo in approximately 100 adults with chronic sleep onset and maintenance insomnia in a parallel group design (the study, with a total N= 223, also included two dosages of estazolam, discussed elsewhere). Subjects were randomized to receive drug or placebo for seven consecutive nights. Outcomes were measured by sleep questionnaires (interval ratings and Likert scales). Dominguez\textsuperscript{149} evaluated a similar population of 45 adults with flurazepam 30 mg or placebo for 7 nights. Sleep variables were assessed by patient questionnaire. Elie\textsuperscript{157} reported on 60 outpatient insomnia patients in a cross-over design such that each patient received a single dose of five different drugs (or drug dosages) or placebo on one night of the week over a five consecutive week period. Study drugs included flurazepam 15 mg, three dosages of loprazolam, and placebo. Outcomes included an index for sleep-onset based on patient questionnaires. Elie\textsuperscript{158} investigated efficacy of flurazepam 30 mg and zopiclone versus placebo over 4 weeks. Flurazepam and placebo groups included 12 chronic insomnia patients per group. Subjects reported sleep outcome variables on post-sleep numerical rating questionnaires. Hartmann\textsuperscript{158} studied 96 adult patients (N=45 for flurazepam + placebo groups) with various insomnia complaints. Subjects were randomly assigned to receive flurazepam 30 mg, secoberbital, l-tryptophan, or placebo for one week of active treatment. Outcomes were assess by sleep logs which included subjective estimates of SL, NOA, duration of awakenings, and QOS.

Mamelak\textsuperscript{160} investigated the effects of flurazepam 30 mg and zopiclone versus placebo in three groups of 10 insomnia subjects per group, each of which received one of the three treatment conditions for 12 consecutive nights. Subjective estimates of SL, TST and NOA were reported. Mamelak\textsuperscript{161} studied 36 elderly patients with chronic insomnia. Each of three groups randomized to flurazepam 15 mg, brotizolam or placebo received drug or placebo for 14 nights. Outcomes included patient-reported SL, NOA, TST and wake time. Daytime performance measures were conducted at the beginning of treatment and following conclusion. Melo de Paula\textsuperscript{162} evaluated flurazepam 30 mg versus placebo and two dosages of lormetazepam in 60 adults with sleep onset or maintenance problems. Subjects received one of the four treatment conditions for two weeks. Outcome data included subjective SL, NOA and TST.

Reeves\textsuperscript{101} investigated the efficacy of flurazepam 15 mg and triazolam versus placebo in 61 geriatric subjects (N=27 for flurazepam + placebo groups) with sleep onset or maintenance insomnia. Subjective sleep outcomes were assessed by interval rating questionnaires. Salkind\textsuperscript{163} evaluated flurazepam 15 and 30 mg versus placebo in 30 general practice insomnia patients. Subjects received each dose of flurazepam and placebo for one week in a crossover trial. Patient-reported SL, TST and QOS were primary outcome variables. Daytime residual effects were also reported. Scharf\textsuperscript{150} studied 243 outpatients (N= 163 for flurazepam vs. placebo) with complaints of sleep initiation or maintenance difficulty. Subjects were randomly assigned to one of three parallel groups – flurazepam 30 mg, estazolam 2 mg or placebo. Treatment condition was administered for 7 nights. Subjects rated sleep latency, TST, QOS and NOA on numerical interval questionnaires. Sunshine\textsuperscript{112} investigated the effects of 15 mg and 30 mg flurazepam versus two dosages of triazolam and placebo in a five-night crossover study, with subjects receiving each intervention for one night. Subjects were 25 inpatients who complained of sleep initiation and maintenance problems. Patients completed sleep questionnaires with interval ratings for TST and NOA.

Kripke\textsuperscript{159} conducted the only identified polysomnographic study of flurazepam. In this study, 99 subjects with chronic insomnia were randomized to one of four parallel groups (flurazepam 15 mg, flurazepam 30 mg, midazolam or placebo). Subjects received treatment for 14 consecutive nights, with PSG recordings on nights 1, 2, 7, 13 and 14. Objective SL, WASO, TST, and SE were reported.
Sleep latency
The only PSG study of 30 mg\textsuperscript{159} found no significant reduction in SL vs. placebo.

Five studies\textsuperscript{148, 150, 160, 162, 163} reported statistically significant improvement (at p <0.05 level) on subjective ratings of sleep onset for flurazepam 30 mg versus placebo. Kripke\textsuperscript{159} found improvement in patient-reported SL for 30 mg only in the early period (nights 1 and 2) of administration. No significant difference from placebo was evident at end of 14-day treatment. Four reports\textsuperscript{112, 143, 149, 158} found no significant subjective improvement in sleep onset with flurazepam 30 mg versus placebo.

Three studies\textsuperscript{101, 157, 163} reported subjectively improved onset at the 15 mg dosage. Kripke\textsuperscript{159} found patient-reported improvement at this dosage only on nights 1-2. Two investigations demonstrated no improvement in sleep onset for flurazepam 15 mg versus placebo.

Total sleep time
Eight studies\textsuperscript{112, 148-150, 159, 160, 162, 163} reported statistically significant improvement for flurazepam 30 mg versus placebo on various subjective scales for sleep duration. One study\textsuperscript{158} reported no significant improvement in duration at this dosage.

Two studies\textsuperscript{112, 163} found significantly improved patient-reported duration at the 15 mg dosage; Kripke\textsuperscript{159} reported subjective improvement only on nights 1 and 2. Likewise, two studies\textsuperscript{101, 161} found no significant subjective improvement in duration for flurazepam 15 mg.

Wake after sleep onset
Two studies reported data for WASO. Kripke\textsuperscript{159} found significantly reduced PSG WASO with flurazepam 30 mg versus placebo. Mamelak\textsuperscript{161} reported no significant reduction in subjective WASO with flurazepam 15 mg in an elderly insomnia population.

Quality of Sleep
Utilizing a variety of differing scales, six studies\textsuperscript{101, 148-150, 157, 163} reported improvement in sleep quality with flurazepam versus placebo. Four studies\textsuperscript{148-150, 163} found improvement at the 30 mg dosage and three studies\textsuperscript{101, 157, 163} at the 15 mg level.

Sleep efficiency
One study\textsuperscript{159} reported PSG sleep efficiency. Flurazepam 30 mg significantly improved efficiency versus placebo.

Number of awakenings
Six studies\textsuperscript{112, 148-150, 158, 160} assessed subjective NOA with flurazepam 30 mg. All found significant reduction in NOA. Three studies\textsuperscript{101, 112, 161} found no significant reduction in NOA with flurazepam 15 mg.

Adverse effects
Cohn\textsuperscript{148} reported that 68% of flurazepam 30 mg subjects experienced an adverse event versus 43% of subjects receiving placebo. Approximately 50% of the flurazepam group reported somnolence, about twice the rate in the placebo population. Dominguez\textsuperscript{149} found a significant increase in side effects for flurazepam 30 mg compared to placebo and stated that 73% of side effects described as “undetermined” were reports of somnolence. Elie\textsuperscript{157} indicated that there was no significant difference in adverse events between flurazepam 15 mg and placebo; likewise Elie\textsuperscript{158} found no difference in rates of somnolence for flurazepam 30 mg versus placebo. Mamelak\textsuperscript{160} found significant performance impairment with flurazepam 30 mg. Mamelak\textsuperscript{161} extended these findings and reported significantly shorter latencies to sleep on MSLT at the beginning and end of treatment. The authors also found significant impairment on digit symbol
substitution and serial learning as well as a significantly slower rate of improvement on reaction, response and movement time. Divided attention was also impaired at end of treatment. Reeves\textsuperscript{101} noted that 6 of 13 flurazepam subjects reported somnolence (versus 4/14 in the placebo group). Salkind\textsuperscript{163} described impaired motor performance in the flurazepam 30 mg group (although not in the 15 mg group) and a significantly higher rate of “hangover effect” at the higher dosage. In the cross-over design, 11 of 30 flurazepam group experienced morning drowsiness/hangover, while this was reported by only 3 of 30 subjects during the flurazepam 15 mg period and 2 of 30 while taking placebo. Finally, Scharf\textsuperscript{150} found AEs in 73% of the flurazepam 30 mg group versus 43% on placebo subjects. Somnolence was the most common event, reported by 57% of flurazepam subjects and 23% of the placebo group.

5.5 Oxazepam

Götestam\textsuperscript{164} studied the efficacy of oxazepam 25 mg vs. placebo with a crossover design in 28 patients with “insomnia.” Subjective reports using interval ratings showed a significant reduction in SL and significant improvement in QOS.

5.6 Quetiapine

One study\textsuperscript{165} investigated the efficacy of quetiapine vs. placebo control in primary insomnia. However, the study included only 13 subjects. Increase in subjective TST and decrease in subjective SL were seen but not statistically significant, given the small sample size.

5.7 Gabapentin

One study\textsuperscript{166} evaluated gabapentin for treatment of primary insomnia. This was an open-label investigation with 18 subjects, variable dosages, and no placebo control. Therefore, the trial was excluded.

5.8 Paroxetine

Two studies assessed paroxetine for treatment of primary insomnia. Nowell\textsuperscript{167} reported a trial of variable dosage in 15 patients, without placebo control. As a result, this investigation was excluded.

Reynolds\textsuperscript{168} evaluated paroxetine 10 mg/20 mg in 27 older adults with primary insomnia who were randomized to drug or placebo. The two doses were pooled for statistical analysis. PSG data showed a modest but significant increase in SL, decrease in WASO and no difference in SE vs. placebo. Sleep quality was improved.

5.9 Trimipramine

Hohagen\textsuperscript{169} studied the effects of trimipramine in 15 adults with primary insomnia. No placebo control was included and, as a result, the study was excluded. Riemann\textsuperscript{170} evaluated 55 adults with primary insomnia in a placebo-controlled double blind study. Dosage was variable (50-200 mg; mean 109.4 mg), but pooled for analysis. No significant difference was observed between trimipramine and placebo for PSG TST or SL but SE was significantly improved with trimipramine. Subjective sleep quality also showed significant improvement.
6.0 DISCUSSION & FUTURE DIRECTIONS

Defining “efficacy”
Assessment of the efficacy of a given agent for the treatment of chronic insomnia is a complex and challenging task. It remains unclear which are the most important variables for defining efficacy. Older studies, particularly the majority of investigations of benzodiazepine efficacy, utilized a variety of predominantly subjective scales and questionnaires. These are highly diverse and did not often include specific numerical patient estimates for sleep outcomes. Since the advent of newer benzodiazepine receptor agonists, more specific and uniform outcomes for both patient-reported and objective (PSG) outcomes (e.g., self-reported and polysomnographic sleep onset latency, wake time after sleep onset, and total sleep time) have been employed, although continued substantial variability in data reporting has not been uncommon.

In addition to the variability in outcome measures reported, there are a number of critical unresolved issues regarding evaluating the efficacy of treatments for chronic insomnia. One is the relative importance of subjective vs objective data. Another is whether metrics of sleep quality, whether they be subjective or objective (e.g. analysis of the microstructure of sleep or related physiological parameters), are perhaps more pertinent than measures of SL, TST or WASO. An additional issue of importance is whether efficacy is better reflected by measures of daytime alertness and cognitive, emotional and psychomotor function than by measures of sleep. Recent behavioral treatment studies in chronic insomnia have taken yet another direction: measuring response or remission of the insomnia syndrome as the most clinically-relevant outcome. This approach makes sense from a patient-centered approach, since most patients complain of “difficulty” falling asleep or staying asleep, rather than tying their complaints to any specific numerical value. Indeed, several studies have identified a group of “non-complaining poor sleepers” whose quantitative sleep measures are similar to those with insomnia. Examining the insomnia syndrome is also useful because it addresses both sleep-related and wake-related symptoms.

Absent clear answers to these questions, the present analysis relies on conventional subjective and objective measures of major sleep variables (sleep onset latency, total sleep time or wake time after sleep onset). The meta-analyses conducted yield recommendations for use of a limited number of drugs for a limited number of specific indications (i.e. sleep initiation and/or sleep maintenance). In all cases, the recommendations are “weak,” in that they are based on relatively limited and low quality evidence. Because of the availability of the requisite type of data, the preponderance of medications included in these analyses are FDA-approved drugs for treating insomnia. This approval is based on FDA standards of significant improvement vs. placebo for one or more indications (i.e. sleep onset or sleep maintenance insomnia). Many agents (including some which are not FDA-approved hypnotics) which we have not recommended for treatment of chronic insomnia have been shown in one or more studies to be “statistically significantly superior” to placebo for a given outcome(s). The discrepancy between 1) FDA approval and demonstration of “statistically significant superiority” to placebo; and 2) the recommendations included in this publication, is clearly a function of our having employed a substantially more demanding systematic approach to evidence quality ratings and establishment of clinical significance thresholds that is not employed in research studies and FDA assessment for approval. The thresholds were determined by clinical judgement of the task force and represent best estimates of the degree of improvement which the “typical patient” would find significant. Although these thresholds are consistent with numerical values that have been recommended as thresholds in contemporary publications, these standards reflect a view of subjectivity, as there are no data which suggest absolute standards for clinical significance. Without question, there may be divergent opinions regarding what constitutes clinical significance and efficacy. This was understood by the task force who assumed that their recommendations are not absolute indications of the presence or absence of clinical utility of a given medication but reflect
their best judgment based on the available data. It is the responsibility of each prescriber to make treatment determinations with this in mind.

Patient selection and inclusion criteria for studies are variable and may substantially impact results for a given outcome (e.g. see Krystal, 2012). Studies not requiring a minimum inclusion criterion for a specific outcome (e.g. inclusion thresholds for SL or WASO) may be underpowered to identify significant change for that outcome.

Understanding the methodology
The recommendations of the TF were developed with the use of GRADE, a state-of-the-art methodology for assessment of clinical data. This approach has distinct strengths, as well as certain limitations. GRADE is a rigorous, detailed and transparent system for evaluation of the relative strengths of evidence for a given intervention. It incorporates numerous considerations which may impact the quality of evidence for a treatment approach. These factors include the heterogeneity of data (i.e. the degree of inconsistency of results across studies), imprecision of the data (i.e. 95% confidence intervals which cross the clinical significance threshold) and potential publication bias (as a result of industry sponsorship). Quality of evidence grades for randomized clinical trials begin at HIGH and are downgraded progressively for heterogeneity, imprecision, and/or potential publication bias. Since the vast majority of studies in this field are industry sponsored, the quality of evidence for nearly all of these studies is, therefore, reduced from HIGH to MODERATE. This is to be expected for clinical trials for many drugs (i.e. not only hypnotics), since the vast majority are, of necessity, industry-sponsored. The extent to which this downgrading of evidence is warranted due to actual publication bias is unknown, but under the GRADE system we have chosen to adopt the conservative approach and assume risk of bias. When heterogeneity and imprecision are accounted for, the quality of evidence for many treatments considered is LOW or VERY LOW. These latter two factors are not uncommon, as there is substantial variability in sleep outcome variables across studies and confidence intervals frequently overlap the clinical thresholds for significance.

Meta-analysis requires specific data (numerical data for a given outcome, presented as mean ± SD). Many studies, particularly older investigations, do not report data in the required format. Some newer publications do not report data in the this format because some sleep variables, particularly sleep onset latency, are not normally distributed. In this case the preferred measure of central tendency is not the mean but the median, the standard deviation may not be a valid measure of the degree of dispersion, and the statistical analyses carried out are not based on the mean and standard deviation. The result of this is exclusion of substantial amounts of data from the formal meta-analyses. While these studies are discussed in the paper and (secondarily) considered in formulation of recommendations, the inability to include such data in meta-analysis represents a distinct limitation.

As described in the methodology section, GRADE requires a recommendation “for” or “against” use of each treatment. When the evidence for efficacy is clear-cut, with 1) relatively high quality of evidence; 2) a high degree of confidence that benefits clearly outweigh harms; and 3) evidence that the effects of treatment are of substantial magnitude, without imposition of significant burden to the patient, a “strong” recommendation is delivered in the form of “we recommend clinicians use X for the treatment of chronic insomnia.” When evidence for benefit is less clear and the quality lower, a “weak” recommendation of “We suggest that clinicians use (or not use)…” However, it is important for clinicians to understand that a recommendation against use, particularly when associated with low quality evidence, is not equivalent to a demonstration of ineffectiveness. Rather, it is often an indication that the available evidence is simply insufficient and fails to provide convincing support in favor of usage by GRADE standards. In the case of drugs (most commonly older drugs) for which none of the data were reported in a form useable for meta-analysis, we refrain from making any recommendation. The specific indications for use of a hypnotic employed in this report are limited to “sleep initiation” and “sleep maintenance.” insomnia. We chose these
since, from a practical clinical consideration, these are the primary complaints with which chronic insomnia patients present and for which clinicians prescribe medication. Moreover, these are the subtypes of insomnia that were actually studied in many investigations — consistent with FDA approval strategies and the matching of drugs to particular types of sleep disturbance.

Hence, some medications may show substantial improvement in TST or sleep quality, yet demonstrate no or insignificant reduction in SL, WASO or NOA to qualify for a recommendation in favor of use.

As described, we established thresholds for clinically significant improvement for each objective and subjective major sleep outcome. Nevertheless, some degree of judgment was introduced in formulating final recommendations. For example, a medication may not have exceeded significance thresholds for both subjective and objective evidence but, when the totality of evidence (including those investigations which could not be included in the meta-analysis) was considered, the TF concluded that a reasonable standard had been met. These considerations also include the role of adverse effects in the decisions made.

Beyond the quality of evidence for or against use of a given drug for sleep onset or maintenance insomnia, the task force also considered the relative benefit:harm ratio and the likelihood that an informed patient would use a specific agent. To a great extent, these decisions are based on clinical judgement. With respect to the benefit:harm consideration, the data on adverse events is often limited or non-existent. In addition, the frequency of some treatment-emergent adverse events (TEAEs) is so low that the reported studies are underpowered to find a difference from placebo. However, this also implies that the effect size for a TEAE would be very small, and hence, it is unlikely that the clinical significance of TEAEs has been underestimated. However, it is worth noting that some TEAEs are very infrequent but very serious when they do occur (e.g., sleep-related behaviors with BzRA). Clinical trials are likely to minimize such risks due to the number of patients treated. As a result of these considerations, assessment of potential harms is largely derived from clinical experience and theoretical considerations, rather than well-documented evidence. This is clearly a limitation of the analysis and further, more systematic investigation of adverse effects is necessary.

As to the issue of likelihood of use by informed patients, the clinical judgement (established prior to formulation of the specific recommendations) is, in most cases, in agreement with recommendations (i.e. an informed patient is likely to use a drug that is recommended and not likely to use one that is not). In certain cases (e.g. see melatonin), the TF considered that, given widespread use and apparently benign side effect profiles, informed patients may be likely to use a specific drug even when data do not clearly support a recommendation for use.

**Clinical application**

Administration of sleep-promoting medication for chronic insomnia is one possible component of what must be a comprehensive approach to evaluation and treatment of chronic insomnia. This approach must include adequate assessment of cause and characteristics of the disorder as well as evaluation and treatment of contributing comorbidities. The latter may include any one or more of numerous medical, neurological and mental disorders, as well as other primary sleep disorders.

Numerous investigations have demonstrated that hypnotic medications are comparably efficacious to cognitive-behavioral therapies (CBT) during active treatment. However, these studies also make clear that the gains associated with CBT are durable following completion of treatment, whereas those associated with medication tend to dissipate following discontinuation of the drug. The vast majority of investigations which are included in the current analysis address relatively short-term use (e.g. one day to five weeks). Some studies have shown that long-term treatment with at least newer generation BzRA hypnotics can be safe and effective under properly controlled conditions. However, chronic use should be
reserved for those individuals for whom CBT is either inaccessible or ineffective, who have been appropriately screened for contraindications to such treatment, who maintain long-term gains with medication, and who are followed regularly. Patient preference must also be considered in the determination of treatment approach.

The investigations which are included in this analysis were focused on “primary” chronic insomnia, with the exception of some older studies (e.g. see zaleplon) which included some patients with “mild” mental disorders. The extent to which these findings apply to chronic insomnia associated with major comorbidities is uncertain, although a limited number of comparative studies suggest at least some degree of efficacy in such cases. It should also be emphasized that the findings presented in this report apply only to adults. None of the agents discussed in this report are approved for use in children and none of the findings presented apply to children or adolescents. There is very little information concerning pharmacotherapy for childhood insomnia. Although independent analyses of efficacy in older adults were not conducted, examination of the findings suggests comparable efficacy across the adult age range. However, it should be emphasized that pharmacokinetics/dynamics in many older adults differ from those of younger adults, necessitating lower starting dosages. The limited information from these studies regarding adverse effects in older adults does not allow meaningful conclusions about the frequency of such events in older patients compared to a younger population.

The data on adverse effects, in general, does not suggest a high frequency of side effects. However, the data on adverse effects is scant and inconsistent, suggesting that caution should be applied in the interpretation of these data. Given the known sedative effects of these agents, particularly those with longer half-lives, clinicians must be diligent in cautioning patients regarding potential complications related to sedation.

**Future Directions**

In an attempt to develop meaningful clinical practice recommendations for the use of sleep-promoting medications, it became increasingly clear to the TF that this endeavor is fraught with multiple limitations. While existing data (especially more recent data) does provide a reasonable foundation for certain recommendations contained in this study, the overall quality of evidence is relatively low in the vast majority of cases. For numerous drugs, there is simply insufficient evidence available to draw on in determining whether or not a compound is efficacious. Data reporting, especially that of older studies, is highly variable and idiosyncratic. As a result, comparing data from one study to another, or conducting meta-analyses of data is not possible. Virtually all studies of prescription hypnotic agents are industry-funded. While the reasons for this are understandable, the potential for publication bias, particularly lack of publication of negative results, compromises the quality of evidence to a significant degree. Moreover, the role of industry in study design and data analysis may further compromise uniformity of data reporting.

With these limitations in mind, the TF recommends the following for future investigations:

1. Clear definitions of inclusion and exclusion criteria
2. Adequately powering studies to detect significant differences for key sleep variables;
3. Development and utilization of uniform data collection instruments which will promote improved cross-study analysis and comparisons;
4. Standardized statistical analysis and data presentation. The majority of newer investigations now present means $\pm$ SD for specific PSG or sleep diary data. For those variables that are not normally distributed, a transformation can be sought which converts the probability distribution to the normal distribution and the transformed mean and SD can then be reported. An effort to report means and SD data should be made for all studies;
5) Although specific numerical data for individual sleep outcome are useful in assessing the efficacy of pharmacological treatment for insomnia, other approaches to such evaluation may be more clinically meaningful. Specifically, determination of the efficacy of a drug in achieving remission of chronic insomnia disorder has been employed in cognitive behavioral treatments for insomnia and should be considered as a clinically relevant outcome in pharmacological trials. This may include not only subjective and objective outcome data for major sleep outcomes, but also sleep quality and daytime functional outcomes;

6) To the extent possible, encourage funding for independent, non-industry investigation of the efficacy of hypnotic medications;

7) Data for adverse events associated with hypnotic medications is collected and analyzed in an even more haphazard way than efficacy data. This is a widespread problem common to studies of all types of medications. Continued efforts should be made to standardize and systematize the reporting of adverse effects data;

8) Daytime sedation, with concomitant risk of motor vehicle or occupational accidents, is a significant potential risk. Further efforts to include objective assessments of performance impairments which may be associated with daytime sedation is encouraged;

9) Virtually no data exists regarding the use of sleep-promoting agents in children. Yet, such medications are not infrequently used in this age group. As such, studies of the efficacy and safety of sleep-promoting medications in children and adolescents should be required.

**Summary**

This analysis is, to the best of our knowledge, the most comprehensive assessment of efficacy of sleep-promoting agents published to date. It relies heavily on rigorous evaluation of the quality of evidence for efficacy, based on GRADE, as well as determination of potential adverse effects, to the extent possible. It is intended to serve as a useful guide for clinicians in prescribing medications for the treatment of chronic insomnia. This analysis, however, also makes it abundantly clear that the availability and quality of the data which serve as the foundation for such recommendations is sorely limited. The result is that many commonly used drugs, including some which carry FDA approval for treatment of insomnia, are not recommended. Further data are required to formulate any reasonable conclusion regarding their efficacy or lack thereof. As a result, clinicians must continue to exercise a reasonable degree of clinical judgment, based not only on these recommendations, but also on clinical experience, prior patient response, patient preferences, and potential adverse effects.
References


59. Ancoli-Israel S, Krystal AD, McCall WV et al. A 12-week, randomized, double-blind, placebo-controlled study evaluating the effect of eszopiclone 2 mg on sleep/wake function in older adults with primary and comorbid insomnia. *Sleep* 2010;33:225-34.


Roth T, Hull SG, Landford DA, Rosenberg R, Scharf MB. Low-dose sublingual zolpidem tartrate is associated with dose-related improvement in sleep onset and duration in insomnia characterized by middle-of-the-night (MOTN) awakenings. Sleep 2008;31:1277-84.


66


121. Krystal AD, Durrence HH, Scharf M et al. Efficacy and Safety of Doxepin 1 mg and 3 mg in a 12-week Sleep Laboratory and Outpatient Trial of Elderly Subjects with Chronic Primary Insomnia. Sleep 2010;33:1553-61.


