労災疾病臨床研究事業費補助金

精神疾患により長期療養する労働者が元の職場等、社会復帰するまでの過程及び
その手法に係る医学的知見や文献等に係る調査研究（160102－01）

精神疾患患者の社会復帰指標作成・効果的介入同定の系統的レビュー (16808287)

平成 28 年度 総括・分担研究報告書

研究代表者　　渡辺 範雄

平成 29 年 (2017 年) 3 月

目次

I. 総括研究報告

労災疾病臨床研究事業費補助金
総括研究報告書

精神疾患患者の社会復帰指標作成・効果的介入同定の系統的レビュー

研究代表者　渡辺　範雄
京都大学大学院医学研究科　健康増進・行動学分野　准教授

研究要旨

精神疾患による長期療養にはなるべく早期の治ゆのみならず職場・社会復帰が望まれ、復帰を予測する指標、また復帰を推進する効果的介入を開発・同定することが社会的急務である。本研究では、国内外の質の高いエビデンスを集積して3つの系統的レビュー(SR)によって精神疾患の疾患重症度や精神状態、睡眠・運動等のライフスタイルデータが社会復帰、また精神療法を利用した、長期療養者の職場復帰予測指標指標作成、これらの結果を統合した予測モデルを作成して、精神疾患による長期療養者の社会復帰推進のための提言を行う。

今年度は先行研究の網羅的収集を行い、現時点での知見と限界点をまとめた。これらの知見から、先行研究よりもより幅広く高感度に文献を収集するための、高感度な文献検索式を作成し、またデータベースもMedline・PubMedに一本化せず幅広く収集するための検索戦略を検討した。さらに、それぞれのSRにて研究計画を作成した。また上記の過程から、精神疾患による休職者の社会復帰に関わる因子には多重共線性があるのを確認し、それを解決しながら定量的に社会復帰の可能性を提示するために、予測モデル作成を研究計画に新たに組み込んだ。

平成29年度は研究計画書に沿ってこれらのレビューを実施し、結果をまとめるとともに英語原著論文化・学会発表など啓発に努める。

A. 研究目的

近年、精神疾患による労災認定請求・支給決定件数は過去最多を記録した。精神疾患で長期療養に至った労働者では、その後の症状の完全消失・社会復帰が望まれるが、薬物療法だけではこれを達成できない。効果的追加介入と、職場・社会復帰の判断指標が必要となる。しかし研究者らの知る限り、良質エビデンスの社会復帰指標は存在しない。

そこで、本研究では、国内外の質の高いエビデンスを集積して3つの系統的レビュー(SR)を行い、疾患重症度、心理社会的因子、ライフスタイル、精神療法介入等が精神疾患による長期療養者の職場・社会復帰にどのような影響を及ぼすのか、明らかにする。またこれらの因子間には多重共線性の存在が強く予測されるため、社会復帰の予測にはこれらの因子を同時に扱う必要がある。そこでこれらの結果および他の文献による結果を統合した予測モデルを作成して、精神疾患による長期療養者の社会復帰推進のための提言を行う。

まず、SR1では観察研究を集積・統合し、精神疾患の疾患重症度やレジリエンス等の精神状態、睡眠の量・質や活動量等、生活指標の縦断的データを利用した、長期療養者の職場復帰予測指標指標

作成を目的とする。

SR2・3では介入研究を集積・統合し、精神疾患による長期療養者の効果的職場復帰促進の介入方法を同定する。具体的には、SR2は無作為割り付け対照試験(RCT)を集積して、症状改善や早期復職のための効果的な心理社会的介入を同定する。またSR3では、RCTに限定せずわが国でも広く行われている復職リハビリテーション等に関する観察研究をも集積し、運動・睡眠等、早期復職のための効果的なライフスタイルに関する介入の同定を行う。

さらに、これらのSRや他の文献の結果から社会復帰に資する因子を同定し、重み付けを行って予測モデルを作成する。

これらの知見により、職場・社会復帰促進に加えてその後の再休職予防や、医療資源の効率的利用に資することができると考えている。

B. 研究方法

まずSR1-3について、各テーマに関する先行研究の網羅的収集を行い、現時点での知見と問題点・限界点をまとめる。それぞれのSRにて研究計画を作成したうえで、先行研究よりもより幅広く高感度に文献を収集するための、高感度な文献検索式を作成する。データベースもMedline・PubMedに一本化せず、幅広く収集するため、PsychINFOやEMBASE、Cochrane CENTRAL等の網羅的なものに関しても検索戦略を立案し、文献検索を行う。

その後研究協力者を含めた2人の独立した研究者による、検索された文献の登録基準該当性確認、データ抽出・研究の質の評価、データのメタ・アナリシスによる統計学的統合、研究の質やアウトカムの重要度を踏まえたうえで可能であればメタ解析を実施する。不可能であっても、GRADEシステムを利用してエビデンスの質・効果サイズの大きさを加味した上で、それぞれ系統的レビューとして推奨を作成する。

また先行研究のレビューにおいて、精神疾患による休職者の社会復帰に関わる因子には多重共線性があるのを確認した。これを解決しながら定量的に社会復帰の可能性を提示するために、SR1-3で同定された因子について、その効果サイズに基づいて重み付けすることで予測モデル式を作成する。これにより、精神疾患によって休職を余儀なくされている休職者が、各時点での精神疾患重症度やライフスタイル、また受療している心理社会的介入や一定のライフスタイル改善の方策を行うことによって、その後の期間毎の復職可能性を定量化する。この結果、今後の復職見込みを各休職者について理想できるようになるだけでなく、各時点での復職を達成するために重点を置くべき介入の推奨を作成することができ、今後の医療施策に資することができる考えている。

それぞれ報告書としてまとめるとともに、英語原著論文・学会等で結果の啓蒙活動を行う。また、職場・社会復帰促進に加えてその後の再休職予防や、医療資源の効率的利用に向けた提言を行う。

（倫理面への配慮）

「人を対象とする医学系研究に関する倫理指針」に基づいて行うが、本研究に含まれる全ての系統的レビュー・予測モデル作成は、一次研究を集積・統合した二次研究であり、不利益・危険性の排除や説明同意、プライバシー保護等は一次研究で担保されているものを扱うため、問題は生じないと考えられる。

C. 研究結果

1. 先行研究の収集

主任研究者・分担研究者間で毎月のオンライン・ミーティングを実施し(平成28年8月12日、9月16日、10月14日、11月18日、12月16日、平成29年1月20日、2月17日、3月10日)、先行研究の網羅的収集を行い、現時点での知見と限界点をまとめた。先行研究ではコホート研究デザインにて社会復帰との関連因子を探索する研究は一定量見られた。しかしながら、これらの質を評価した上で統合した量的系統的レビューは見られなかった。

心理社会的介入に関しては、うつ病・適応障害等の軽症精神障害、統合失調症等の重症精神障害に関する文献レビューを行った。それぞれの疾患に関して疾患特異的な精神療法のみならず、復職を目的とした精神療法を行っている無作為割り付け対照試験(RCT)や、それらを統合した系統的レビューを同定した。しかし、これらは主に欧米のデータのみであり、日本のリワーク等の復職プログラムは事例

検討レベルでのエビデンスが多数を占めていた。また欧米でも、特に統合失調症を中心とした臨床上の効果が注目されている認知機能改善療法の優れた系統的レビューはなく、本研究の枠組みで取り組むことを確認した。

2. 本研究の計画策定

これらの知見から、先行研究よりもより幅広く高感度に文献を収集するための、高感度な文献検索式を作成し、またデータベースもMedline・PubMedに一本化せず幅広く収集するための検索戦略を検討した。さらに、それぞれのSRにて研究計画を作成し、英文化も進めている。

また上記の過程から、精神疾患による休職者の社会復帰に関わる因子には多重共線性があるのを確認し、それを解決しながら定量的に社会復帰の可能性を提示するために、予測モデル作成を研究計画に組み込んだ。

3. データベースの高感度文献検索式

国際標準となっている研究デザインごとの高感度フィルター等を組み込み、また幅広く網羅的に文献を収集するための文献検索式を作成した。データベースもMedline・PubMedのみならず、PsycINFOや国内のデータベースである医学中央雑誌の利用等を盛り込んだ。

進捗として、多少の研究計画の変更はあったものの、概ね計画通りに進行している。平成29年度には計画書に沿ってレビューを実際に作成し、その結果を統合して予測モデルを確立する。

D. 考察

本研究では、精神疾患による長期療養者の職場復帰予測指標の同定と職場復帰促進の効果的介入方法同定という両側面を、強度の低い運動や睡眠等ライフスタイル・レベルと、強度の高い疾患特異性やレジリエンスを踏まえた医療レベルに分けてレビューすることで、職場復帰の促進や、医療資源の効率的利用に資する。さらにこれらの因子を独立ではなく、重み付けして社会復帰予測モデルを提案することで、複合的因子を総合した社会復帰指標が開発される。

本研究は、系統的レビュー国際標準であるコクラン方式を踏襲して行う。研究者らは全員皆それぞれ系統的レビュー作成経験があるため研究遂行可能性は高いと考えている。

E. 結論

本研究は、精神疾患による長期療養者の社会復帰予測指標の開発と、社会復帰促進の効果的介入方法同定とを文献の系統的レビューにより行う。これにより、職場・社会復帰促進に加えてその後の再休職予防や、医療資源の効率的利用に資することができると考えている。

また、本研究ではこれらレビューの結果をもとに、社会復帰のための予測モデルを作成する。文献レビューをもとにしたこの試みは、筆者らの知る限り世界でも例をみないため、新たな知見が得られる可能性がある。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

1. Kusaka K, Shinohara K, Tada M, Watanabe N, Furukawa TA. Concerns About Selective Outcome Reporting. J Clin Oncol. 2017;35(6):688.
2. Itani O, Jike M, Watanabe N, Kaneita Y. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. Sleep Med. 2017;32:246-256.
3. Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, Twisk J, DeRubeis RJ, Dimidjian S, Hegerl U, Mergl R, Jarrett RB, Vittengl JR, Watanabe N, Cuijpers P. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br J Psychiatry. 2017;210(3):190-196.
4. Furukawa TA, Schramm E, Weitz ES, Salanti G, Efthimiou O, Michalak J, Watanabe N, Cipriani

A, Keller MB, Kocsis JH, Klein DN, Cuijpers P. Cognitive-Behavioural Analysis System of Psychotherapy (CBASP), a drug, or their combination: differential therapeutics for persistent depressive disorder: a study protocol of an individual participant data network meta-analysis. BMJ Open. 2016;6(5):e011769.
5. Furukawa TA, Salanti G, Atkinson LZ, Leucht S, Ruhe HG, Turner EH, Chaimani A, Ogawa Y, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Suganuma A, Watanabe N, Stockton S, Geddes JR, Cipriani A. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. BMJ Open. 2016;6(7):e010919.
6. De Crescenzo F, Foti F, Ciabattini M, Giovane DC, Watanabe N, Schepisi MS, Quested DJ, Cipriani A, Barbui C, Amato L. Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis [protocol]. Cochrane Database Syst Rev. 2016.

2. 学会発表
なし
（発表誌名巻号・頁・発行年等も記入）

H. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得
なし

2. 実用新案登録
なし

3. その他
なし

II. 分担研究報告

労災疾病臨床研究事業費補助金
分担研究報告書

労働者における精神疾患による休業からの復職に関連するライフスタイルを含む予測因子についての体系的レビュー

研究分担者　井谷　修
大分大学医学部　公衆衛生・疫学講座　准教授

研究要旨

労働者の病気による長期休業の原因として精神疾患が占める割合が増加しており，精神疾患に罹患し休業した労働者に対してReturn To Work（RTW）を促進させる施策を検討する必要がある。更に，施策の検討にあたって，精神疾患の症状改善・治癒といった疾患そのものの回復過程でどのような要因がRTWに関連しているのかを明らかにすることが急務である。そこで本研究では精神疾患による休業からのRTWに関連する要因についての体系的レビューの実施を計画した。

今年度は，先行研究について収集し，現時点での知見と限界点をまとめた。そして，先行研究よりもより幅広く高感度に文献を収集するための、高感度な文献検索式を作成し、またデータベースもMedline・PubMedに一本化せず幅広く収集するための検索戦略を検討した。また，研究の質の評価方法についても最新知見に基づき検討を行った。

今後は，研究プロトコルを確立した上で，体系的レビューの本実施を行う予定である。これにより，精神疾病の休業者の復職を検討する上での有用な基礎的データが提供可能である。

A. 研究目的

先進国において労働者の精神疾患罹患による経済的および社会的な損失が大きな社会問題となっている。精神疾患に対応するための費用は，国家総予算の約3～4%に達するという試算もある。これらの費用のうち約3分の1が精神疾患への直接の医療費であるが，残りの3分の2は労働者における生産性の低下や雇用の喪失によるものである。特に，ここ数十年の間で，労働者の病気による長期休業の原因として精神疾患が占める割合は増加してきている。このような状況の改善のため，精神疾患に罹患し休業した労働者に対してreturn to work（RTW）を促進させる施策を検討する必要がある。施策の検討にあたって，精神疾患の症状改善・治癒といった疾患そのものの回復過程でどのような要因がRTWに関連しているのかを明らかにすることが急務である。

疾病への対応を検討する際に，その要因を科学的に分析することは，対応を効率的に進める上で非常に重要である。一般的に認識されていることとして，病気による休業の要因は多因子的であり，単に原因となっている疾病のみではない。労働者の疾病による休業の先行研究として，現在まで身体疾患や怪我に関するものについては，休業の要因についての研究が数多く行われ，それらのエビデンスを集約した体系的レビューも行われてきている。しかし，精神疾患による休業の予後を左右する要因ついての先行研究は少ない。そこで本研究では精神疾患による休業から

のRTWに関連する要因についての体系的レビューの実施を計画した。

B. 研究方法

まず，本研究に関連する先行研究の網羅的収集を行い，現時点での知見と限界点をまとめた。そして，研究計画を作成したうえで，先行研究よりもより幅広く高感度に文献を収集するための、高感度な文献検索式を作成する。データベースもMedline・PubMedに一本化せず，幅広く収集するため，PsychINFOやEMBASE，Cochrane CENTRAL等の網羅的なものに関しても検索戦略を立案し，文献検索を行う。

その後2人の独立した研究者により，検索された文献の登録基準該当性確認・データ抽出・研究の質の評価・データのメタ・アナリシスによる統計学的統合を行う。

（倫理面への配慮）

「人を対象とする医学系研究に関する倫理指針」に基づいて行うが、本研究に含まれる全ての系統的レビュー・予測モデル作成は、一次研究を集積・統合した二次研究であり、不利益・危険性の排除や説明同意、プライバシー保護等は一次研究で担保されているものを扱うため、問題は生じないと考えられる。

C. 研究結果

1. 先行研究の収集

先行研究については，Blank et al.は精神疾患とRTWまたは長期休業に関連する要因についての体系的レビューを行っている。その結果，14編の研究を抽出し， RTWに関連した要因として家族歴やhealth risk behaviorsや社会状況や精神疾患の病状といったものを指摘している。 しかし，このレビューでは6ヶ月以上休業を扱った研究は対象から除外していることや，対象とする研究デザインとして横断研究を含むすべての研究タイプを対象としているため，同定した関連要因とRTWとの関連について因果関係を述べることが出来ないなど，無視できない大きなlimitationが存在していたといえる。また，Cornelius et al.が縦断研究のみを対象とした系統的レビューを行っているが，この際の抽出論文は7編のみであり，要因の同定に充分な研究数を確保できていない。

2. 本研究の計画策定

本研究では，先行研究のlimitationを克服するべく，より包括的な検索を行うことと，より先進的な方法を用いた検索・bias評価を行うことを検討した。なお，本研究では，観察研究をレビュー対象とする。

まず，本研究の臨床疑問・仮説の設定(PICOS)として

- Participants: 精神疾患による休業者
- Intervention: 生活習慣全般（短睡眠時間・不眠症者・運動習慣なし・食事悪い・メディア使用過多・休養無し），重度精神疾患，レジリエンス弱い等）
- Comparison: Iの逆
- Outcome:復職 return to work
- Study design: 全ての観察研究

と設定した。

文献検索においては，精神疾病について扱った論文を抽出するためのフィルタおよび観察研究を抽出するためのフィルタについてそれぞれ最新のものを使用し，更にRTWについては可能な限りのイディオムを用いた上で網羅的な検索が可能となる検索式を作成した。またデータベースもMedline・PubMedに一本化せず幅広く収集するための検索戦略を検討した。

また，集積する個々の研究の質評価に用いるツールについても最新のものを使用するべく，調査を行った。その結果，非ランダム化試験用のバイアスのリスク評価ツールであるRoBANS: The Risk of Bias Assessment tool for Non-randomized Studiesを用いて評価することとした。

D. 考察

本研究の特徴は，第一に可能な限り包括的に精神疾病によるRTWに関連する要因について集積を行うことである。特に，対象とする要因については，睡眠や食事や運動やメディア使用といったライフスタイル全般を包括的に対象に含めることとした。第二に，対照研究タイプを前向き研究に限定することを検討している。これにより要因の因果関係を論じることが可能となり，高いエビデンスレベルの提供が可能である。

E. 結論

今後は研究の本実施に向けて，研究プロトコルの確定を行い，計画した系統的レビューを実施する予定である。この研究により，職場・社会復帰促進に加えてその後の再休職予防や，医療資源の効率的利用の検討に有用な基礎データを提供できると思われる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

1. Itani O, Kaneita Y: The association between shift work and health: a review. Sleep and Biological Rhythms 14:231-239, 2016.
2. Itani O, Kaneita Y, Munezawa T, Ikeda M, Osaki Y, Higuchi S, Kanda H, Nakagome S, Suzuki K, Ohida T: Anger and Impulsivity among Japanese adolescents: A nationwide representative survey. J Clin Psychiatry 77:e860-866, 2016
3. Itani O, Kaneita Y, Munezawa T, Mishima K, Jike M, Nakagome S, Tokiya M, Ohida T: Nationwide Epidemiological Study of Insomnia in Japan. Sleep Med. 25:130-138, 2016.
4. Morioka H, Itani O, Osaki Y, Higuchi S, Jike M, Kaneita Y, Kanda H, Nakagome S, Ohida T: Association between Smoking and Problematic Internet Use among Japanese Adolescents: Large-Scale Nationwide Epidemiological Study. Cyberpsychol Behav Soc Netw. 19:557-561, 2016.
5. Mihara S, Osaki Y, Nakayama H, Sakuma H, Ikeda M, Itani O, Kaneita Y, Kanda H, Ohida T, Higuchi S. Internet use and problematic Internet use among adolescents in Japan: A nationwide representative survey. Addictive Behaviors Reports. 4: 58-64, 2016.
6. Morioka H, Itani O, Osaki Y, Higuchi S, Jike M, Kaneita Y, Kanda H, Nakagome S, Ohida T: The association between alcohol use and problematic Internet: A large-scale nationwide cross-sectional study on adolescents in Japan. J epidemiol. 27:107-111, 2017.
7. Itani O, Jike M, Watanabe N, Kaneita Y: Short sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression. Sleep Med. (*in press*).
8. Kinjo A, Imamoto A, Ikeda M, Itani O, Ohida T, Kaneita Y, Kanda H, Tanihata T, Higuchi S, Osaki Y. The Association Between Alcohol-Flavoured Non-Alcoholic Beverages and Alcohol Use in Japanese Adolescents. Alcohol Alcohol. (*in press*).

2. 学会発表

なし

H. 知的財産権の出願・登録状況 （予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

労災疾病臨床研究事業費補助金
分担研究報告書

精神疾患による長期療養者の職場復帰に関する、症状改善・固定や早期復職のための心理社会的介入

研究分担者　小川 雄右
京都大学大学院医学研究科　健康増進・行動学分野　助教

研究要旨

精神疾患は現在、長期休職の最も大きな原因であり、精神疾患に罹患した就労者の休職期間を短くし、疾病罹患前の生産性を回復することは、患者本人のみならず家族、職場、社会にとって重要である。本研究では精神疾患による長期療養者に対する職場復帰のための心理社会的介入に関する系統的レビューを実施し、効果的な介入方法の検討を行う。

本年度は先行研究の検索、同定を行った。それぞれの疾患に対する心理社会的介入の効果を検討されている無作為割り付け対照試験(RCT)や系統的レビューを同定した。うつ病、適応障害、統合失調症に関しては、様々な心理社会的介入の効果を検討したRCT及び系統的レビュー・メタアナリシスが報告されていた。しかし、自閉症スペクトラム障害については、文献レビューは存在するものの、包含されたRCTは包括的ではなく、またメタアナリシスも行われていなかった。先行研究からの知見、限界を検討し、新しい系統的レビューの研究計画を作成した。

A. 研究目的

精神疾患は現在、長期休職の最も大きな原因であり(Kessler 2003)、精神疾患により休職することは貧困や社会的孤立にも関連する (Henderson 2011)ことが指摘されている。休職期間を短くし、疾病罹患前の生産性を回復することは、患者本人のみならず家族、職場、社会にとって重要である。これらの達成のためには、薬物療法だけではなく、心理社会的介入が重要であることが指摘されている。

そこで、本研究では国内外の質の高いエビデンスを集積して系統的レビューを実施し、精神疾患による長期療養者に対する効果的な心理社会的介入方法の検討を行う。具体的には、これまでに実施された無作為割り付け対照試験(RCT)を同定し、症状改善や早期復職のための効果的な心理社会的介入を同定する。本研究により、職場復帰に加えて再休職の予防や生産性の回復、医療資源の効率的利用に資することができると考えられる。

B. 研究方法

本研究ではコクラン共同計画の方法により、系統的レビュー、メタアナリシスを実施する。

まず精神疾患による長期療養者の職場復帰に関する、症状改善・固定や早期復職のための心理社会的介入に関する先行研究の収集を実施し、これまで

の知見と限界をまとめた。

本研究で対象とするのは、精神疾患による長期療養者に対する心理社会的な介入の効果を検討している全ての RCT である。先行研究よりも幅広く研究を同定するために、感度の高いデータベースの検索式を作成する。検索は MEDLINE に限定せず、CENTRAL や PsychINFO、Scopus のデータベース、ClinicalTrials.gov などの臨床研究登録データベースにおいて実施し、更には本研究のテーマに関連する系統的レビュー、本研究で同定した文献が引用した文献、本研究で同定した文献を引用した文献の調査も行い、関連研究の包括的な調査を行う。さらに、解析の直前には最新の RCT のアップデートサーチも実施する。

本研究で検討する介入は、認知行動療法・対人関係療法・問題解決療法などの精神療法、リラクセーション、復職支援プログラム、認知機能改善療法、その他、復職を目的にした全ての心理社会的介入を検討する。個人に対する介入であるか、グループに対する介入であるかは問わない。アウトカムは復職、仕事能率、疾患重症度、業務に関するコミュニケーションスキルなどである。

2人の研究者が独立して、検索によって得られた文献が包含基準を満たしているかを確認し、データ抽出、研究の質の評価、結果の統計学的統合を実施し、GRADEシステムによる推奨を作成する。

（倫理面への配慮）

本研究は既存の資料を用いた二次研究であるため、研究対象者に対する人権擁護上の配慮、研究対象者に対する不利益、危険性の排除、説明と同意（インフォームド・コンセント）に関わる状況などに関する問題は生じず、倫理面での問題は生じないと判断した。

C. 研究結果

本年度は、先行研究の検索、同定を行い、現時点での知見と限界をまとめた。うつ病、適応障害、統合失調症、自閉症スペクトラム障害など、休職、職場復帰に関わることの多い精神障害に関する文献レビューを実施し、それぞれの疾患に関して心理社会的介入の効果を検討されているRCTや、それらを統合した系統的レビューを同定した。うつ病、適応障害、統合失調症に関しては、様々な心理社会的介入の効果を検討したRCT及び系統的レビュー・メタアナリシスが実施されていた。しかし、自閉症スペクトラム障害については、文献レビューは存在するものの、包含されたRCTは包括的ではなく、またメタアナリシスも行われていなかった。

先行研究からの知見、限界を検討し、新しい系統的レビューの研究計画を作成した。また、本研究では先行研究よりも網羅的に文献を収集するために、データベースはMEDLINEに限定せず、CENTRALやScopus、PsycINFOその他のデータベースも検索を行い、より広く文献を同定できる検索式を作成し、幅広く研究を同定するための戦略を検討した。

D. 考察

本研究では、精神疾患による長期療養者の職場復帰促進の効果的介入方法を、系統的レビュー、メタアナリシスを実施することにより検討する。先行研究からの知見、限界を検討し、本研究では先行研究よりも網羅的に文献を収集するために、データベースはMEDLINEに限定せず、CENTRALやScopus、PsycINFOその他のデータベースも検索を行い、より広く文献を同定できる検索式を作成し、幅広く研究を同定するための戦略を検討した。また、自閉症スペクトラム障害については、我々の知る限りでは、職場復帰のための介入を検討したRCTの包括的な系統的レビュー及びメタアナリシスは存在しないため、新たな知見が得られる可能性がある。

E. 結論

本研究では精神疾患による長期療養者の職場復帰促進の効果を検討したRCTの系統的レビューを実施することにより、精神疾患による長期療養者の社会復帰促進のための効果的介入方法の同定を行う。これにより、職場復帰促進に加えてその後の再休職予防や生産性の回復、医療資源の効率的利用に資することができると考えられる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

1. Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, Hayasaka Y, Chaimani A, Salanti G. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. Lancet Psychiatry. 2016. 3(11):1059-1066.
2. Furukawa TA, Salanti G, Atkinson LZ, Leucht S, Ruhe HG, Turner EH, Chaimani A, Ogawa Y, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Suganuma A, Watanabe N, Stockton S, Geddes JR, Cipriani A. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. BMJ Open. 2016. 6(7):e010919

2. 学会発表

なし

H. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

労災疾病臨床研究事業費補助金
分担研究報告書

精神疾患による長期療養者の職場復帰に関する、運動・睡眠等、ライフスタイルに関する介入

研究分担者　地家　真紀
日本大学医学部社会医学系公衆衛生学分野　助教

研究要旨

精神疾患等により休職する労働者は増加傾向にあり、休職者の復職支援は行政、事業者、労働者にとって重要な課題となっている。復職支援プログラムは複数あるが、本研究課題では運動・睡眠等ライフスタイルを含めた復職支援プログラムに着目し日本及び諸外国での先行研究を把握することを目的としてMedline・PubMed及び医中誌を用いて文献の検索を行った。

Medline・PubMedでの検索において精神疾患による休業者を対象としたシステマティックレビューが存在した。しかしながら集積された論文の一部に薬物療法の介入が含まれているものもあり、本研究課題とは異なることが明らかにされた。

医中誌での検索において、集団認知行動療法を実施した介入研究が確認できた。一方でプログラムの効果について対象者のプログラム実施前、実施後を評価したもの、企業や医療機関での取り組みの報告が多数を占めた。現時点で質的量的に担保された研究報告が少ないことが明らかになった。

来年度は『対象者：精神疾患による長期療養者、介入：運動、睡眠等ライフスタイルについてのプログラムを含めた非薬物療法の実施、結果の評価：復職までの日数、再休職率』にて検索を実行する。日本国内で実施されているプログラムは日本語の報告が多いことから医中誌を用いて日本語文献の集積をする。ROBANSで研究の質の評価、GRADEで推奨を作り国際標準に応じたレビューを実施する。

A. 研究目的

「業務上疾病発生状況等調査」や「過労死等の労災補償状況」において「強い心理的負荷を伴う業務による精神障害」、「精神障害に関する事案」の件数は増加傾向にあり、職域における精神障害は公衆衛生学的な課題となっている[1,2]。

このような状況に対し厚生労働省は平成18年に「労働者の心の健康の保持増進のための指針」を公表した[3]。さらに平成27年12月からストレスチェックを実施し[4]、労働者個人および集団のストレス状況の分析評価することにより職域におけるメンタルヘルス対策が講じられるようになった。

その一方で厚生労働省が実施した「平成25年　労働安全衛生調査」によるとメンタルヘルス対策の取り組みを行っている事業場の割合は60.7％であり、金融業・保険業では91.8％と高値を示す一方で、生活関連サービス業・娯楽業は45.8％と産業によって状況

はさまざまであることが明らかにされている[5]。また独立行政法人 労働政策研究・研修機構が実施した「メンタルヘルス、私傷病などの治療と職業生活の両立支援に関する調査」によると通常の年次有給休暇以外で連続して休職できる病気休職制度（慣行含む）をもつ事業場は調査対象の91.1％であること[6]、さらに休暇の期限は事業場により異なり1年以下（「3か月まで」9.6％、「3か月超から6か月まで」13.3％「6か月超から1年まで」22.3％）と規定する事業場が45.2％であることが報告された。このような状況を踏まえ休職者は期日までに勤務可能なレベルまで症状の改善をしなければ退職せざるを得ない。つまり復職支援は今後の人生にかかわる大きなサポートであるといっても過言ではない。

休職者に対して実施される復職支援について日本国内外から報告されている。例えばオランダの適応障害と診断を受けた対象者に問題解決するための介入を行ったところ、通常ケアの群と比較して休職期間が短かった[7]。またScheneらは通常ケアに加えて作業療法を行った群において休職期間が短く、仕事のストレスの増加は認められなかったと報告した[8]。我が国においては休職者にリワークプログラムが示され、行政、事業場、労働者それぞれの立場から取り組まれている。リワークプログラムは3段階から構成される複合的なトレーニングであり対象者の状態に応じて復職を目指す支援として知られる[9,10]。復職支援のプログラムは医療機関で実施される医療リワークをはじめ複数のプログラムが存在するが、本研究課題では運動・睡眠等ライフスタイルに焦点をあてたプログラムについて検討する。

日本で実施されている研究、取り組みは日本語で報告されていることが多いため医中誌を用いた検索を実行する。さらに諸外国で実施された英語文献と対比することにより日本の復職支援の現状を把握することを目的に調査を実施した。

B. 研究方法

精神疾患により療養、職場復帰を望む者を対象にライフスタイルについての介入研究や復職支援についての論文をMedline・PubMed及び医中誌を用いて網羅的に検索を行った。

（倫理面への配慮）

「人を対象とする医学系研究に関する倫理指針」に基づいて研究を実施したが、本研究は文献を系統的に検索しレビューを行うため個人の特定は不可能であり、対象者個人への不利益はないと考える。

C. 研究結果

1．Medline・PubMedによる検索

精神疾患による休業者を対象としたシステマティックレビューが存在した。Joyceらは、精神疾患者の職場での介入についてシステマティックメタレビューを実施していた[11]。1990年1月以降の文献を対象にAMSTARという評価方法を用いて評価を行っている。身体活動について従業員のメンタルヘルスに影響する可能性はあるが、必要な活動のタイプ、量、強度は不明であると示された。また、Nigatuらが報告したシステマティックレビューでは精神疾患休職者の復職促進の介入の効果について集積、評価がされた[12]。運動のみならず復職促進を目的としたさまざまな介入方法をまとめてレビューを行っている。

2. 医中誌による検索

日本における精神疾患療養者の復職について、原著論文の検索を実行した。プログラムの効果について対象者のプログラム実施前、実施後を評価したもの、企業や医療機関での取り組みの報告が多数を占めた。介入研究は2報確認できたが、一報は休業終了後に復職した労働者を対象とした文献であったため本研究課題とは異なった[13]。もう一報は研究協力者を募集しランダムに介入群、通常診療群に割り付けプレセッション1回及び本セッション8回の集団認知行動療法が実施された。介入群において抑うつ状態、非機能的認知、社会的問題解決能力が有意に改善しうつ病の休職者の復職支援に認知行動療法は有効であることが報告された[14]。

D. 考察

精神疾患による休職者の復職支援は日本のみならず諸外国でも実施されていることが確認された。そして本研究課題である運動・睡眠等ライフスタイルを含んだ介入についてシステマティックレビューの先行研究が報告された[7,8]。本研究課題である運動はさまざ

まな種類がある。身体活動と出勤しているのに心身の不調により生産性が低下してしまう状況や職場の福祉に関する先行研究においても介入に用いられた運動は複数でありJoyceらが示した通り運動はその種類、実施する時間や頻度についても評価する必要があるのかもしれない。

日本において復職支援のひとつにリワークプログラムがある。これは3つの段階から構成され、第1段階として生活リズムや症状の回復、第2段階として疾病の理解や発症要因の分析、第3段階として対人関係能力の改善が行われるものである[9]。また医療機関が実施する精神科治療、再休職予防を目的とした医療リワーク、障害者職業センターが実施する支援プランに基づく支援を行う職リハリワーク、企業内、従業員支援プログラム（EAP）などで実施される労働をさせてよいか見極める職場リワークが存在しさまざまな機関で復職支援が実施されている[10]。これら復職支援プログラムについて原著論文が多く発表されているが介入効果について介入の前後で評価された論文が多く客観的評価に乏しい状況である。またランダム化比較試験を実施した論文は一報確認できた[14]。質的、量的に十分な研究が少ないことから運動・睡眠等ライフスタイルだけではなく非薬物療法の範疇までカバーをした検索式を作成し、検討することが必要だと考える。

来年度は『対象者：精神疾患による長期療養者、介入：運動、睡眠等ライフスタイルについてのプログラムを含めた非薬物療法の実施、結果の評価：復職までの日数、再休職率』にて検索を実行する。また日本国内で実施されているプログラムは日本語の報告が多いことから医中誌を用いて日本語文献の集積をする。ROBANSで研究の質の評価、GRADEで推奨を作り国際標準に応じたレビューを実施する。

E. 結論

運動・睡眠等ライフスタイルについての介入を実施した研究は存在したが、日本人を対象としたものは少なく網羅的な検索が求められる。来年度は医中誌で検索を行い、さらにMedline・PubMed等を用いて諸外国の研究と比較を行う。

参考文献

1. 業務上疾病発生状況等調査 厚生労働省 http://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei11/h26.html
2. 過労死等の労災補償状況 厚生労働省 http://www.mhlw.go.jp/file/04-Houdouhappyou-11402000-Roudoukijunkyokuroudouhoshoubu-Hoshouka/h27_seishin.pdf
3. 労働者の心の健康の保持増進のための指針 厚生労働省 http://www.mhlw.go.jp/new-info/kobetu/roudou/gyousei/anzen/dl/101004-3.pdf
4. 厚生労働省 改正労働安全衛生法に基づくストレスチェック制度について http://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/pdf/150422-1.pdf
5. 厚生労働省 平成 25 年 労働安全衛生調査 http://www.mhlw.go.jp/toukei/list/dl/h25-46-50_01.pdf
6. メンタルヘルス、私傷病などの治療と職業生活の両立支援に関する調査 http://www.jil.go.jp/institute/research/2013/documents/0112.pdf
7. van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ. Reducing long term sickness absence by an activating intervention in adjustment disorders: a cluster randomised controlled design. Occup Environ Med 2003;60:429–437.
8. Schene AH, Koeter MW, Kikkert MJ, Swinkels JA, McCrone P. Adjuvant occupational therapy for work-related major depression works randomized trial including economic evaluation. Psychol Med. 2007 ;37(3):351-62.
9. 五十嵐良雄 休職・失職している人はどのくらいいるの？ 安全と健康 2016;17(3):77－9
10. 五十嵐良雄 3つのリワーク、その違い 安全と健康 2016;17(4):80－9
11. Joyce S, Modini M, Christensen H, Mykletun A, Bryant R, Mitchell PB, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med. 2016; 46: 683-97.
12. Nigatu YT, Liu Y, Uppal M, McKinney S, Rao S, Gillis K, et al. Interventions for enhancing return to work in individuals with a common mental illness: systematic review and meta-analysis of randomized controlled trials. Psychol Med. 2016; 46: 3263-74.
13. 昇淳一郎，石田裕美，小林慎治，木村映善，石原謙 復職支援プログラム「超短時間日々面談」実施による精神症状改善効果 松仁会医学誌 2013; 51(2)112－8.
14. 田島美幸、中村聡美、岡田佳詠。うつ病休職者のための集団認知行動療法の効果の検証 産業医

学ジャーナル 33(1), 54-59, 2010-01
15. Brown HE, Gilson ND, Burton NW, Brown WJ. Does physical activity impact on presenteeism and other indicators of workplace well-being? Sports Med. 2011 1;41(3):249-6.

F. 健康危険情報

なし

G. 研究発表

1. 論文発表
Itani O, Jike M, Watanabe N, Kaneita Y. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. Sleep Med. In press.

2. 学会発表
なし

H. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得
なし

2. 実用新案登録
なし

3. その他
なし

III. 研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

書籍

なし

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Itani O, Jike M, Watanabe N, Kaneita Y	Short sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression.	***Sleep Med.***	32	246-256	2017
Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, Twisk J, DeRubeis RJ, Dimidjian S, Hegerl U, Mergl R, Jarrett RB, Vittengl JR, Watanabe N, Cuijpers P.	Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials	***Br J Psychiatry***	210	190-196	2017
De Crescenzo F, Foti F, Ciabattini M, Giovane DC, Watanabe N, Schepisi MS, Quested DJ, Cipriani A, Barbui C, Amato L	Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis [protocol].	***Cochrane Database Syst Rev***		CD012364	2016
Itani O, Kaneita Y	The association between shift work and health: a review.	***Sleep and Biological Rhythms***	14 (3)	231-239	2016
Itani O, Kaneita Y, Munezawa T, Mishima K, Jike M, Nakagome S, Tokiya M, Ohida T	Nationwide Epidemiological Study of Insomnia in Japan.	***Sleep Med.***	25	130-138	2016

Morioka H, Itani O, Osaki Y, Higuchi S, Jike M, Kaneita Y, Kanda H, Nakagome S, Ohida T	The association between alcohol use and problematic Internet: A large-scale nationwide cross-sectional study on adolescents in Japan.	***J epidemiol.***	27 (3)	107-111	2017
Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, Hayasaka Y, Chaimani A, Salanti G.	Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies.	***Lancet Psychiatry***	3(11)	1059-1066.	2016

IV. 研究成果の刊行物・別刷

Sleep Medicine 32 (2017) 246–256

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression

Osamu Itani [a, 1], Maki Jike [b, 1], Norio Watanabe [c, *], Yoshitaka Kaneita [a]

[a] Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, Oita, Japan
[b] Division of Public Health, Department of Social Medicine, Nihon University School of Medicine, Tokyo, Japan
[c] Department of Health Promotion and Human Behavior, Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form
1 August 2016
Accepted 15 August 2016
Available online 26 August 2016

Keywords:
Diabetes mellitus
Hypertension
Mortality
Meta-regression
Sleep deprivation
Vascular disease

ABSTRACT

Objective: The dose–response of short sleep duration in mortality has been studied, in addition to the incidences of notable health complications and diseases such as diabetes mellitus, hypertension, cardiovascular diseases, stroke, coronary heart diseases, obesity, depression, and dyslipidemia.
Methods: We collected data from prospective cohort studies with follow-ups of one year or more on associations between short sleep duration and the outcomes. For the independent variable, we divided participants at baseline into short sleepers and normal sleepers. The primary outcomes were defined as mortality and an incident of each health outcome in the long-term follow-up. Risk ratios (RRs) for each outcome were calculated through meta-analyses of adjusted data from individual studies. Sub-group and meta-regression analyses were performed to investigate the association between each outcome and the duration of short sleep.
Results: Data from a cumulative total of 5,172,710 participants were collected from 153 studies. Short sleep was significantly associated with the mortality outcome (RR, 1.12; 95% CI, 1.08–1.16). Similar significant results were observed in diabetes mellitus (1.37, 1.22–1.53), hypertension (1.17, 1.09–1.26), cardiovascular diseases (1.16, 1.10–1.23), coronary heart diseases (1.26, 1.15–1.38), and obesity (1.38, 1.25–1.53). There was no sufficient usable evidence for meta-analyses in depression and dyslipidemia. Meta-regression analyses found a linear association between a statistically significant increase in mortality and sleep duration at less than six hours. No dose–response was identified in the other outcomes.
Conclusions: Based on our findings, future studies should examine the effectiveness of psychosocial interventions to improve sleep on reducing these health outcomes in general community settings.

1. Introduction

Short sleepers are prevalent throughout the world. In the U.S., the age-adjusted mean sleep duration was 7.18 hours and the prevalence of sleepers reporting less than six hours of sleep was 29.2% in 2012 [1]. In several developed countries, the prevalence rate is not higher than in the U.S., but 11.3% in Canada and 9.8% in the U.K. [2].

Several systematic reviews have shown that short sleep duration is associated with important health outcomes including not only mortality [3–6] but also hypertension [5,7], cardiovascular diseases [8], stroke [9], diabetes mellitus [10,11], and obesity [12]. These have been regarded as phenotypes of metabolic abnormalities [13] or arteriosclerosis promotion [14,15] associated with short sleep duration. However, because these reviews investigated associations between short sleep duration and these health outcomes utilized various methodologies in conducting reviews, another systematic review may be needed where the same methodology is used across health outcomes. In this review, associations between short sleep and incidents of some important health outcomes, including dyslipidemia and depression, which have not yet been examined in previous reviews, should also be investigated. Although sleep duration less than six hours is reported to be

Abbreviations: CER, control event rate; CI, confidence interval; HR, hazard ratio; MOOSE, meta-analysis of observational studies in epidemiology; NOS, Newcastle–Ottawa scale; OR, odds ratio; PRISMA, preferred reporting items for systematic reviews and meta-analyses; RR, risk ratio.

Previous presentation: Poster presented at the 73rd Annual Meeting of Japanese Society of Public Health, Tochigi, Japan, November 5–7, 2014.

* Corresponding author. Department of Health Promotion and Human Behavior, Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. Fax: +81-75-753-4641.
E-mail address: noriowncu@gmail.com (N. Watanabe).

[1] OI and MI contributed equally to this work as the first authors.

http://dx.doi.org/10.1016/j.sleep.2016.08.006
1389-9457/

associated with higher risk than that of seven to eight hours (especially in terms of mortality outcome in previous cohort studies [16–19]), to the best of our knowledge, this has not been systematically examined in meta-analyses and meta-regression, which can contribute to publication bias and to let researchers speculate on mediator effects of sleep duration on health outcomes.

We therefore conducted a systematic review, meta-analyses, and meta-regression to examine if short sleep duration is associated with a higher prevalence of health outcomes using the same.

2. Methods

We performed the study in accordance with the PRISMA (preferred reporting items for systematic reviews and meta-analyses) [20] and the MOOSE (meta-analysis of observational studies in epidemiology) [21] guidelines, with these checklists (see Appendices S5 and S6 in the Supplementary material).

Two independent researchers (OI and MJ) separately assessed the eligibility, extracted data, and checked the quality of the included studies. Any disagreements were resolved through discussion between these two, and with a third reviewer (NW) if disagreements persisted.

2.1. Data sources and searches

The studies were initially identified on October 17, 2013, through a search of PubMed, PsycINFO, CINAHL, and Embase using pre-specified search terms (Appendices S1–S4). Major medical journals, conference proceedings, and reference lists of included studies and previous systematic reviews were also hand-searched for published, unpublished, and ongoing studies. To identify new studies published during the review process, we conducted a search of PubMed using the same search strategy on October 9, 2014 and on May 6, 2016.

2.2. Study selection

We included studies with a prospective cohort or randomized controlled trial design, conducted in community settings, which compared short with normal sleepers for mortality and incidence of health outcomes in a long-term follow-up. We limited studies to those with a minimum follow-up duration of one year from baseline, and a minimum of 20 participants. Studies were excluded if most participants were aged 20 years or younger at baseline, or if participants had been diagnosed with the health outcome at baseline. We also excluded studies that were conducted in inpatient settings and those that involved pharmacological interventions.

The eligibility of each study for inclusion was checked at two stages: (1) looking through the title and abstract and (2) checking the full text.

2.3. Data extraction and quality assessment

2.3.1. Definition of sleep duration

The definition of short sleep was based on the original paper because common sleep duration varies among cultures and ethnicities [22,23]. Durations of short sleep were incorporated into subgroup analyses and meta-regression as mediators (see below). When both a subjective (eg, sleep diary) and objective sleep duration (eg, actigraphy or polysomnography) were reported, we selected the former as the independent variable. Although a self-report survey may be unlikely to capture the actual amount of sleep per night in comparison with actigraphy [24] or polysomnography [25], objective measures may not always be utilized in general community settings and subjective measures might be preferable because of their applicability. When both sleep durations per day (possibly including a daytime nap) and per night were reported, we selected the latter.

The duration of normal sleep was also defined based on the original paper.

2.3.2. Outcome measures

The outcome was defined as mortality and incidence of health outcomes, which were diabetes mellitus, hypertension, dyslipidemia (hypo or hyperlipidemia), cardiovascular diseases (including events in the heart and brain), coronary heart diseases, stroke, obesity, and depression. When a formal diagnosis was not provided, a surrogate outcome (eg, coronary artery calcification instead of diagnosis of coronary artery diseases, a self-report of diabetes mellitus without evidence of formal diagnosis) was included in the primary analyses, but a sensitivity analysis was planned (see below).

2.3.3. Assessment of bias

We employed the Newcastle–Ottawa scale (NOS) [26] to assess the studies' quality. The instrument has three broad categories (patient selection, four criteria, comparability of study groups, one criterion, and assessment of the outcome, three criteria). For the comparability criteria, we allotted two stars according to the depth of statistical adjustment for risk factors in the original studies (eg, one star for age, sex, and race only, two stars for beyond these). Therefore, a study could reach a full mark with nine stars. For the second and third items of the outcome criteria, we defined, a priori, follow-up durations as reasonably long enough, and adequate follow-up of cohorts in terms of the percent lost to follow-up that was allowed for each disorder (ie, three years and 10% for any cause of mortality, two years and 20% for diabetes mellitus, two years and 20% hypertension, two years and 20% for dyslipidemia, three years and 10% for cardiovascular diseases, three years and 10% for coronary heart diseases, two years and 20% for obesity, and two years and 20% for depression, respectively).

Although previous meta-analyses [27,28] deemed quality of a study as high when it had five or more stars on the NOS criteria, we (a priori) set eight or more stars as high in order to focus on very high quality studies.

2.4. Data synthesis and analysis

We analyzed data y and conducted a meta-analysis for each dependent outcome. In the meta-analysis, we calculated risk ratios (RRs) by pooling adjusted RRs between short and normal sleep provided by the original studies with a random effects model. If hazard ratios (HRs) were reported in a study but RRs were not, the HRs were regarded as RRs. Among studies where odds ratios (ORs) were provided but not RRs, we calculated RRs by using the ORs and control event rates (CERs) in normal sleepers reported in the original studies. Regarding studies where both RRs and CERs were not reported, and only ORs were provided, CERs were borrowed from a study whose characteristics were similar. In the primary analyses, regarding studies where RRs were provided for subgroups separately (eg, male and female), data from these subgroups were combined using a fixed-effect meta-analysis.

Statistical heterogeneity between studies was investigated using the I^2 statistic [29], assuming an I^2 of 75% or greater to be an important level of inconsistency, as a previous review employed [30]. To assess publication bias, we used a funnel plot and Egger's test for all primary outcomes [31]. We used the "trim and fill" method to adjust the funnel plot and recalculated the results [32].

Although subgroup analyses should be interpreted with caution [33], we planned, a priori, to perform analyses for several types of baseline characteristics (ie, between 20 and 65 years of age, or aged 65 years or more; male or female).

Sensitivity analyses were planned for the primary outcome by: (1) excluding studies with surrogate outcomes, and limiting studies to those (2) where sleep duration was reported per night, (3) with eight or more stars in the NOS, (4) those following up participants for no less than ten years, and (5) excluding studies where CERs were borrowed from the others to calculate RRs. Moreover, for the purpose of exploring mediator effects of the duration of definition of short sleep, subgroup analyses were performed for clustering included studies according to the duration (eg, less than five hours or six hours) for each outcome. When ten or more studies were included [34] and studies were clustered into three or more levels in each outcome, meta-regression analyses were also performed to examine a linear association between sleep duration and increase in the outcome frequency, using a random-effects model and illustrating, if any, regression line and its 95% prediction intervals.

A P-value of less than 0.05 was chosen to test the null hypothesis despite multiple comparisons in order to avoid type II over type I errors. For all outcomes, 95% confidence intervals (CIs) were calculated. The data were analyzed using the Comprehensive Meta-Analysis Software (Version 3) [35], with double data entry to avoid input errors.

3. Results

3.1. Search results

The initial electronic search yielded 3580 articles, and an additional database search identified 182 studies on October 9, 2014 and 388 on May 6, 2016. In total, 2521 studies remained after removing duplicate articles. A hand-search did not identify any studies that had not been included in the electronic search (Fig. 1).

At the first and second eligibility check stages, two independent researchers identified 277 articles with an inter-rater reliability of

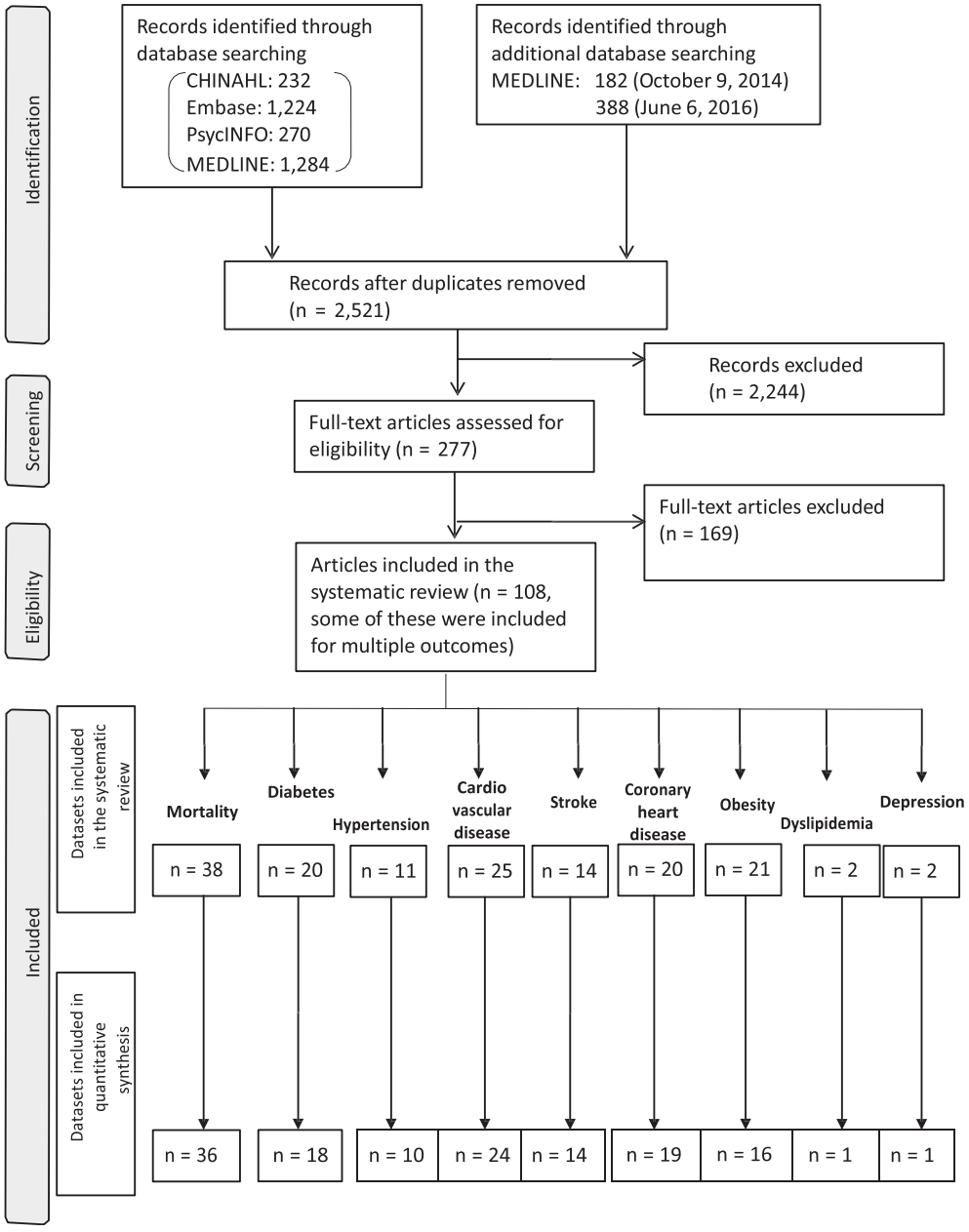


Fig. 1. Flowchart for the included studies.

kappa statistic at 0.79 (95% CI: 0.75–0.82) and 108 articles at 0.64 (0.54–0.73), respectively.

3.2. Characteristics of included studies

All of the 108 included studies were prospective cohort studies. Forty-five studies reported data for two or more health outcomes, and 153 datasets for nine outcomes (N = 5,172,718) were collected from the 108 studies. Most studies were conducted in developed countries (see Table 1 and Tables S1–S8 in the Supplementary material). The number of participants in each dataset ranged from 276 to 392,164, the duration of follow-up was from one to 30 years, and the total NOS scores were from five to nine. The studies' definitions of short sleep varied, but most defined it as less than five or six hours.

Of the included studies for the systematic review, we could not pool data from 14 datasets in the meta-analyses, mainly because no usable data for meta-analyses were provided (Table S9). The number of datasets included in the meta-analyses on each outcome varied from ten (hypertension) to 36 (mortality). Table 1 shows the characteristics of the studies included for the mortality outcome.

3.3. Effect estimates of short sleep compared to normal sleep from meta-analyses

3.3.1. Primary analyses

Compared with normal sleep, short sleep showed a statistically significant increase in mortality due to all causes at a RR of 1.12 (95% CI 1.08–1.16, $P < 0.005$, $I^2 = 25\%$, N of datasets = 36; Figs 2 and 3). Qualitatively similar significant results were obtained for diabetes mellitus, hypertension, cardiovascular disease, coronary heart disease, and obesity, at RR of 1.37 (1.22–1.53, $P < 0.005$, $I^2 = 53\%$, N = 18), 1.17 (1.09–1.26, $P < 0.005$, $I^2 = 48\%$, N = 10), 1.16 (1.10–1.23, $P < 0.005$, $I^2 = 49\%$, N = 24), 1.26 (1.15–1.37, $P < 0.005$, $I^2 = 51\%$, N = 19), and 1.38 (1.25–1.53, $P < 0.005$, $I^2 = 60\%$, N = 16), respectively (Fig. 2). In terms of stroke, short sleep showed no statistically significant increase compared to normal sleep at a RR of 1.08 (0.98–1.19, P = 0.10, N = 14). No important level of heterogeneity between datasets was observed for all outcomes. For the depression and the dyslipidemia outcomes, only one study was included in each and meta-analyses were therefore not performed.

3.3.2. Possible publication bias for primary analyses

In the funnel plots and results from Egger's test (see Figs S1–S3 in the Supplementary material), significant publication bias was observed for coronary heart disease (P = 0.005) and obesity (P = 0.037). However, all the imputed RRs using the trim and fill method still indicated a statistically significant increase in incident of these outcomes among short sleepers.

3.3.3. Subgroup analyses for age groups

Subgroup analyses for those aged 65 years or more and those under 65 at baseline were conducted (Fig. 2). For outcomes where both age groups were analyzed, short sleep was associated with a significant increase in incidents of outcome among those aged under 65 years but did not among those aged over 65 years for hypertension, cardiovascular disease, and obesity, in comparison with normal sleep. Short sleep was not associated with a significant increase in either age group for the mortality outcome; however, short sleep was associated significantly in both age groups for the coronary heart disease outcome. For outcomes where only subgroup analyses were conducted for those aged less than 65 years, short sleep was associated with a significant increase in incidents of diabetes mellitus, but not for stroke.

3.3.4. Subgroup analyses for sex

Subgroup analyses for sex were performed for outcomes other than depression and dyslipidemia. In comparison with normal sleep, short sleep was associated with a significant increase in both sexes in mortality, cardiovascular disease, coronary heart disease, and obesity. Short sleep was associated with a significant increase among males but not females for diabetes mellitus, but vice versa for hypertension and stroke (Fig. 2).

3.3.5. Sensitivity analyses

A majority of sensitivity analyses showed qualitatively similar results to those in the primary analyses (Fig. 2), although the result was not statistically significant in an analysis limiting studies judged to be high quality using the NOS in hypertension outcome (RR = 0.93, 95% CI 0.47–1.84, N = 2) (P = 0.835).

3.3.6. Subgroup analyses and meta-regression for the duration of definition of short sleep

Subgroup analyses for the duration were conducted for the outcomes other than depression and dyslipidemia (Fig. 4). In comparison with normal sleep, short sleep defined as the duration less than six hours was associated with a significant increase in mortality, diabetes, cardiovascular disease, and obesity. In hypertension, stroke, and coronary heart disease outcomes, this trend was not observed probably due to a small number of studies in a cluster.

Meta-regression analyses were performed for mortality, diabetes, hypertension, cardiovascular, and obesity outcomes. In mortality outcome, shorter duration of definition of short sleep was significantly associated with increase in the outcome (coefficient = −0.056, SE = 0.021, P = 0.008, R^2 analog = 0.84, Fig. 5), and the 95% CIs of prediction curves were above zero at the definition of sleep less than six hours. In obesity outcomes, however, shorter sleep duration was associated with decrease in the outcome (coefficient = 0.181, SE = 0.089, P = 0.042, R^2 analog = 0.33, Fig. S4), and the 95% CIs of prediction curves were above zero at the definition of sleep more than five hours. No significant association was observed in diabetes (P = 0.237), hypertension (P = 0.655), or cardiovascular disease (P = 0.819) outcomes.

4. Discussion

To the best of our knowledge, this is the first systematic review where the association between short sleep duration and multiple important health outcomes was investigated with the same methodology used in meta-analyses, and where dose–response of short sleep duration on these outcomes was explored in meta-regression analyses. The present review revealed that short sleepers were likely to be associated with greater mortality than normal sleepers with a RR of around 1.12, in other words around a 12% absolute increase. For the other outcomes, short sleepers were likely to have a point estimate of an absolute increase of 37% for diabetes mellitus, 17% for hypertension, 16% for cardiovascular disease, 26% for coronary heart disease and 38% for obesity. In terms of depression and dyslipidemia, no sufficient evidence from meta-analyses existed to conclude whether short sleep was associated with an increase in the incidents in the meta-analyses.

The present results are similar to those of previous systematic reviews, such as a RR of 1.12 (N of studies = 15) for the mortality outcome [4] and of 1.23 (N = 6) for the hypertension outcome [5]. We believe that the findings of our review add to those in these previous reviews because of our updated comprehensive search and rigorous methodology. However, further epidemiological studies are needed to investigate the association of short sleep with dyslipidemia and depression.

Table 1
Characteristics of studies for the mortality outcome.

Author, Year	Sample size	Male %	Mean age ± SD (range) in years at baseline	Years of follow up Mean ± SD	Definition of Short sleep duration (h)	Definition of normal sleep duration (h)	Newcastle–Ottawa Scale: Selection/Comparability/Outcome
Tsubono et al. 1993 [36]	4,318	39.8	61.4 (40+)	4	≤6/night	7–8/night	★★☆★/★★/★★★
Kojima et al. 2000 [37]	5,322	45.8	Male: 46.9 (20–67) Female: 47.7 (20–67)	11.9	−6.9/night	7.0–8.9/night	☆★☆★/★★/★★★
Seki et al. 2001 [38]	1,065	41.3	65.3 ± 3.6 (60–74)	7.5	<6/day	7/day	★★☆★/★★/★★★
Heslop et al. 2002 [39]	Base line: 7,028 2nd screening: 3,030	Baseline: 85.7 2nd screening: 85.4	Male: (−65) Female: (−60)	25	<7/day	7–8/day	★★☆★/★★/★★★
Mallon et al. 2002 [40]	1,870	48.4	56 (45–65)	12	<6/night	6–8/night	★★☆★/★☆/★★★
Goto et al. 2003 [41]	724	34.7	Male: median 73 (65–97) Female: median 74 (65–97)	12	<6/day	6–7/day	★★★★/★★/★★★
Amagai et al. 2004 [18]	11,325	39.0	55.1 (19–93)	8.2 ± 1.5	≤5.9/night	7.0–7.9/night	★★★★/★★/★★★
Patel et al. 2004 [42]	82,969	0.0	53.4 (40–65)	14	≤5/day	7/day	☆★☆★/★★/★★★
Tamakoshi et al. 2004 [19]	104,010	42.2	56.6 (40–79)	9.9	≤4/day	7/day	★★☆☆/★★/★★★
Ferrie et al. 2007 [43]	Phase1: 9,781 Phase3: 7,729		(35–55)	Phase1: 17.1 Phase3: 11.8	≤5/night	7/night	☆★☆★/★★/★★★
Lan et al. 2007 [44]	3,079	56.8	Male: 71.3 (64+) Female: 71.9 (64+)	8.4 ± 3.3	<7/night	7–7.9/night	★★★★/★★/★★★
Gangwisch et al. 2008 [45]	9,789	37.2	Male: 45.0 (32–59) Female: 73.0 (60–86)	−8, −10	≤5/night	7/night	★★★★/★★/★★★
Ikehara et al. 2009 [46]	98,634	42.1	Male: 58.8 (40–79) Female: 60.2 (40–79)	Median 14.3	≤4/day	7/day	★★☆★/★★/★★★
Mallon et al. 2009 [47]	3,523	49.7	40 ± 10 (30–65)	20	<6/night	≥6/night	★★☆★/★★/★★★
Stone et al. 2009 [48]	8,101	0.0	77.0 (69+)	6.9	<6/night	6–8/night	★★★★/★★/★★★
Suzuki et al. 2009 [49]	12,601	51.1	74.1 ± 5.4 (65–85)	5.3	≤5/day	7/day	☆★☆★/★★/★★★
Chien et al. 2010 [50]	3,430	47.3	(35+)	15.9 (13.1–16.9)	≤5/day	7/day	★★☆★/★★/★★★
Mesas et al. 2010 [51]	3,820	43.8	71.8 ± 7.9 (60+)	6.8	≤5/day	7/day	★★☆★/★★/★★★
Vgontzas et al. 2010 [52]	1,741	42.6	Male: 50.1 ± 14.5 (20+) Female: 47.4 ± 12.6 (20+)	Male: 14 Female: 10	<6/night without insomnia, <6/night with insomnia	≥6/night without insomnia	★★★★/★★/★★★
Castro-Costa et al. 2011 [53]	1,512	38.3	68.9 ± 7.1 (63–75)	7.5 Median: 8.9	<6/night	7 to <8/night	★★★★/★★/★★★
Kripke et al. 2011 [54]	434	0.0	67.6 ± 7.9 (50–81)	10.5	<5/day	6–6.5/day	★★★★/☆☆/★★★
Kronholm et al. 2011 [55]	23,290	48.8	(25–64)	29–34	≤5/night	7–8/night	★★☆★/★★/★★★
Cohen-Mansfield et al. 2012 [56]	1,166	55.5	83.4 ± 5.3 (75–94)	20	<7/night	7–9/night	★★☆★/★★/★★★
Ensrud et al. 2012 [57]	2,505	100.0	75.7 ± 5.2 (67+)	3.4 ± 0.5	≤5/day	>5/day	☆★★★/★★/★★★
Chen et al. 2013 [58]	4,064	55.8	73.8 ± 5.7 (65+)	9	≤4/night	7/night	★★★★/★★/★★★
Garade et al. 2013 [59]	4,941	100.0	(40–59)	30	<6/day	6–7/day	☆★★★/★★/★★★
Hale et al. 2013 [60]	3,942	0.0	62.1 (50–79)	11–16	≤5/night	7–8/night	★★☆★/★★/★★★
Kakizaki et al. 2013 [61]	49,256	48.2	(40–79)	10.8	≤6/day	7/day	★★☆★/★★/★★★
Kim et al. 2013 [62]	135,685	45.6	(45–75)	12.9	≤5/day	7/day	★★☆★/★★/★★★
Li et al. 2013 [63]	9,455	(38.1)	(20–79)	7	≤5/night	7/night	★★☆★/★★/★★★
Magee et al. 2013 [64]	227,815	46.3	(45+)	2.8	<6/day	7/day	★★☆★/★★/★☆★
Yeo et al. 2013 [65]	13,164	41.4	(20+)	9.44	≤5/day	7/day	★★★★/★★/★★★
Bellavia et al. 2014 [66]	70,973	53.3	(45–83)	15	<6/day	6.6–7.4/day	★★☆★/★★/★★★
Rod et al. 2014 [67]	9,098	67.2	45 (35–55)	22	≤5/night	7/night	☆★☆★/★★/★★★
Xiao 2014 [68]	239,896	56.2	(51–72)	14	<5/night	7–8/night	★★☆★/★★/★★☆
Zawisza 2015 [69]	2,449	35.0	Male: 72 ± 5.8 Female: 2.5 ± 5.7	22	≤5/day		★★★★/★★/★★★
Zuurbier 2015 [70]	1,734	46.6	62.2 ± 9.3 (45–98)	7.3 ± 1.3	<6/night	6–7.5/night	☆★★★/★★/★★★
Hall 2015 [71]	3,013	48.6	73.6 ± 2.9 (70–79)	8.2 ± 2.3	<6/night	7–8/night	★★★★/★★/★★★
Cai 2015 [72]	113,138	60.1	Male: (40–75) Female: (44–79)	Median: Male: 6.07 Female: 7.12	4–5/day	7/day	★★★★/★★/★★★

Filled stars indicate applicable for the item. The empty stars indicate not applicable for the item.

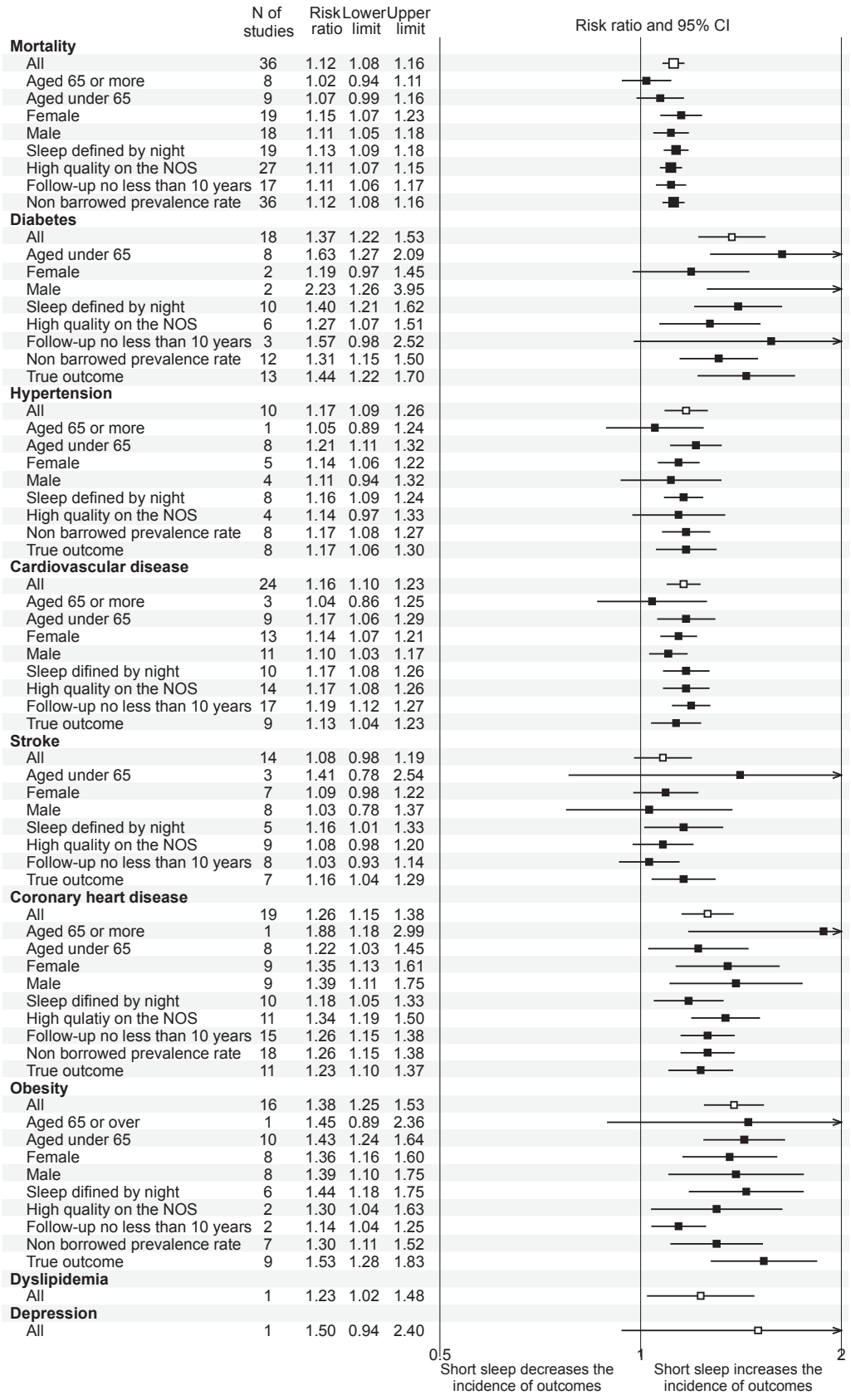


Fig. 2. Relative risks of mortality and health outcomes comparing short with normal sleepers. NOS, Newcastle–Ottawa Scale. For the outcomes of dyslipidemia and depression, no meta-analyses were performed.

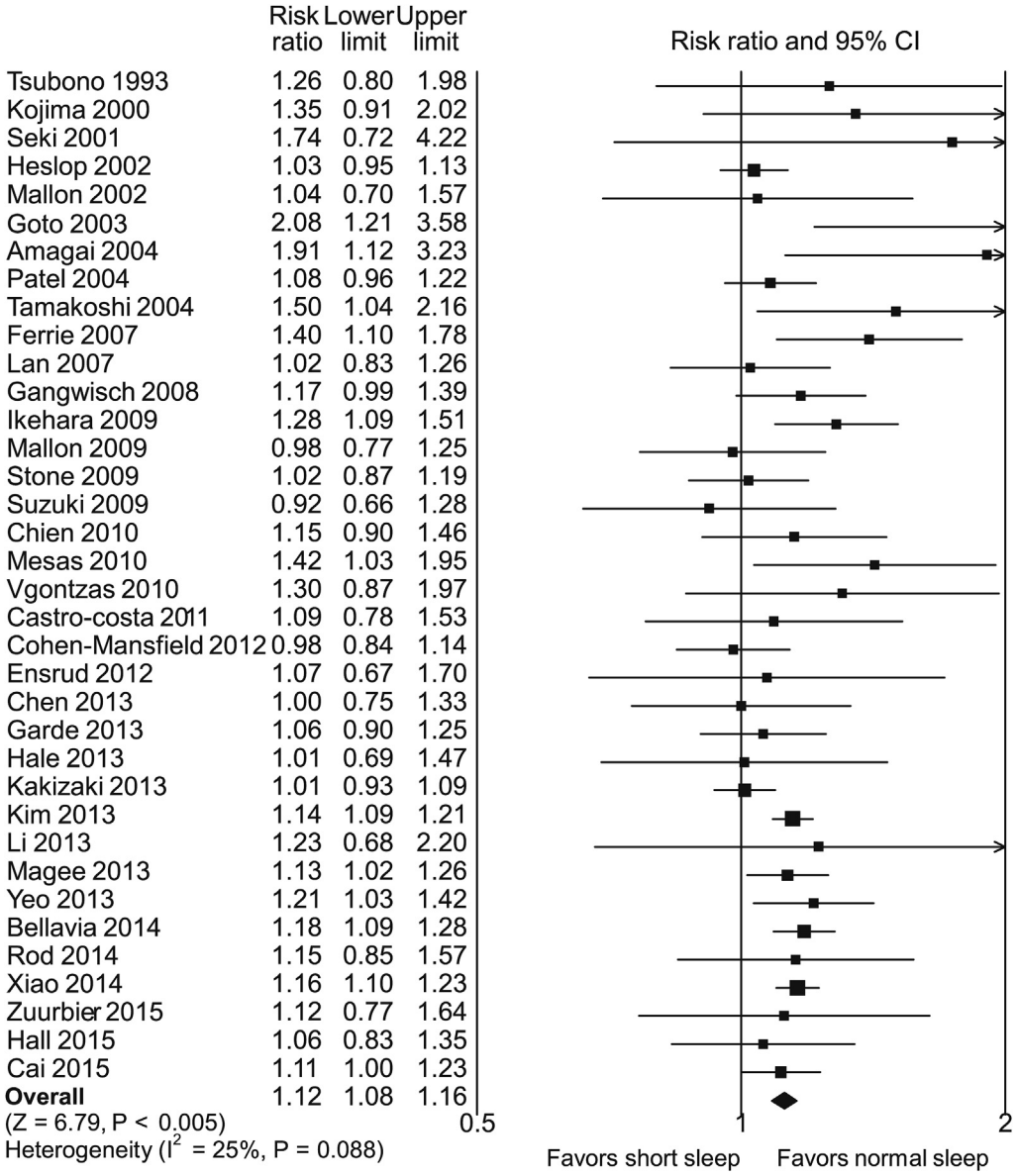


Fig. 3. All-cause mortality outcome comparing short with normal sleepers.

Our study (as in a previous review [5]) reported that in subgroup analyses, a gender difference was found in associations between short sleep duration and health outcomes, such as diabetes mellitus and hypertension. However, each gender may have one illness but not the other in association with short sleep, as observed in the present review.

Subgroup analyses also showed that short sleep defined as a duration less than six hours was associated with a significant increase in mortality, diabetes, cardiovascular disease, and obesity. In meta-regression, shorter duration of sleep definition was linearly associated with an increase in the mortality outcome. A significant association in the opposite trend was observed in the obesity outcome, and no significantly linear association was observed for the other outcomes; however, we were not able to conclude whether the association existed or not because of the small numbers of studies included in each outcome. An important limitation of prior epidemiological studies is that definitions of short sleep across studies can lead to prevention of adequate comparisons [73]. We believe this issue can be approached by the methodology we used in the present review.

Although the study confirmed that short sleep duration was associated with an increase in the important health outcomes, the mechanisms do not seem straightforward. Previous studies have showed that sleep deprivation reduces energy expenditure in healthy men [74], an insulin-resistant state in human adipocytes [75], and inadequate pancreatic insulin secretion and increased plasma glucose concentrations after a meal [76]. Thus, chronic sleep curtailment may lead to consequent outcomes related to metabolic systems. In addition, short sleep duration has been reported to be associated with increased ghrelin and decreased leptin, which is likely to increase appetite [77,78], possibly explaining increase in obesity and diabetes mellitus [79]. As well as changes in these hormones, sympathetic nervous system overactivity and changes in circadian rhythm may play an important role in development of hypertension [80,81]. Cardiovascular diseases can be caused by complex combinations of these metabolic and circulatory abnormalities. Although these endocrine and cardiovascular illnesses had been adjusted in statistical model in most of the studies investigating the mortality outcome, these illnesses might have occurred or worsened during the course in participants and caused

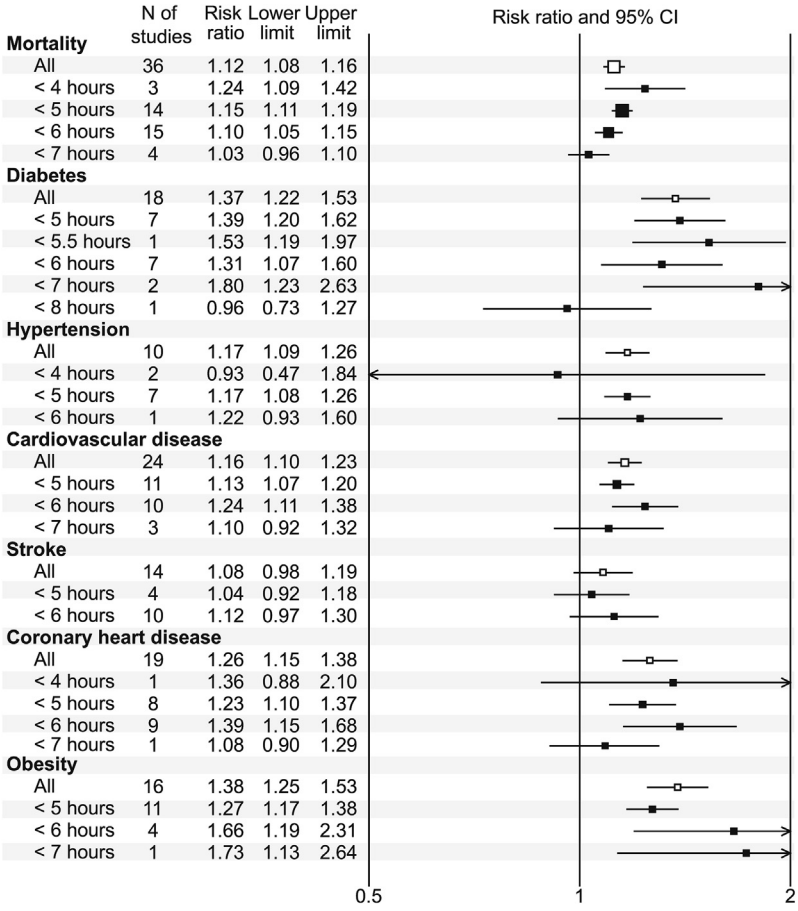


Fig. 4. Subgroup analyses for the duration of the definition of short sleep.

subsequent death. Although our findings appear to be very convincing, one may need to practice caution in terms of considering implications of the results in the community. First, although short sleep is associated with increased mortality and other health outcomes, there is no rigorous evidence that lengthening sleep duration can lead to smaller frequency of these outcomes. We do not intend to recommend hypnotics for people whose sleep is less than six hours a night. With regard to psychoeducation, psychotherapy, and psychosocial interventions, many studies have focused on insomnia and the efficacy of these interventions for the quality of sleep, which was reported in recent systematic reviews [82,83]. To the best of our knowledge, there has been no evidence to recommend these interventions for the community for the purpose of decreasing the frequency of mortality and other health outcomes. It should also be noted that the role of individual differences regarding sleep duration preferences is still uncertain [84]. Unmarried persons [85], those with lower socioeconomic status, or those who engage in frequent binge drinking [86] have been reported to be more likely to be short sleepers. It seems likely that psychoeducation or psychosocial intervention about the present review's results of the relationship between short sleep and health outcomes can encourage people to modify their behaviors; yet, the effects of this intervention may not lead to subsequent decrease in the frequency of health outcomes, and their preferences in terms of sleep should be taken into consideration. A pragmatic effectiveness comparative study in general community settings is needed to determine if psychoeducation, psychosocial, or lifestyle interventions can lead to prevention of premature death and the other important health outcomes.

4.1. Limitations of the study

First, we did not investigate the relationship between long sleep duration and mortality or health outcomes. Long sleep duration has been reported to be more widespread than short sleep [2]. Although a critical review argues that it is premature to conclude a U-shaped association between sleep duration and mortality risks [87], previous systematic reviews have repeatedly suggested that long sleep can be associated with an increase in mortality [3,4,6,73,88,89], hypertension [5,90,91], diabetes mellitus [91,92], and cardiovascular diseases [8,92]. We decided not to report outcomes associated with long sleep duration because of the journal's limited space, but we are planning (July 2016) a systematic review of long sleep duration using the same methodology.

Impacts of the quality of sleep in either subjective or objective measures on the prevalence of the health outcomes were also not investigated in the present review. Although previous studies have reported the association between insufficient sleep such as insomnia and mortality [63,93], we intended to focus on the duration of sleep because it is likely that sleep duration is more easily recognized and accurately reported by participants rather than the quality of sleep in community surveys. However, interventions to reduce mortality and the incidents of the other health outcomes may include information about not only sleep duration but also sleep quality. Previous systematic reviews on cognitive behavioral therapy for insomnia (CBT-I) delivered through the Internet [83] and self-help interventions [94] have revealed that CBT-I is likely to improve the sleep quality, thus components of these low-intensity interventions might be

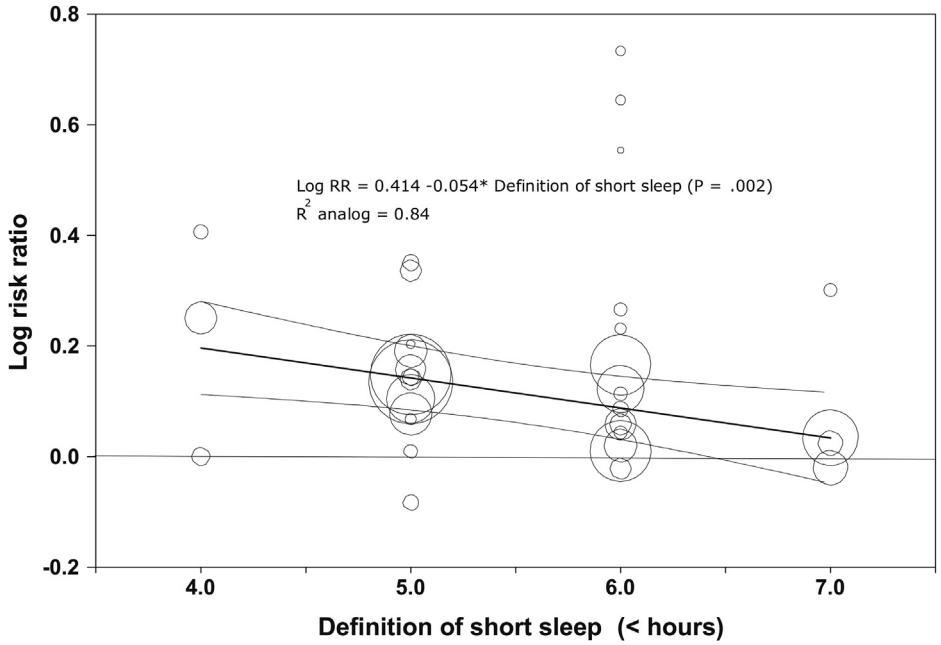


Fig. 5. Meta-regression for the duration of the definition of short sleep in mortality. Regression line and its prediction intervals (95% CIs) were presented in the figure.

considered to be incorporated to those conducted in the community settings.

Third, we did not include previous important surveys [15,69,95] that seemed eligible for the present systematic review, because we focused on studies that defined short and normal sleep duration and compared the incidence of health outcomes between those within these two conditions. These studies reported odds ratios of outcomes for each hour increase in sleep duration, and we regretted that we were not able to include the results from the study. However, we intended to include as many studies as we could by using several methodologies, such as calculating a RR from an OR using a CER. We believe that this is among merits of our systematic review.

5. Conclusions

Short sleep defined as the duration less than six hours was associated with a significant increase in mortality, diabetes, cardiovascular disease, coronary heart disease, and obesity. For clinical practice, the effectiveness of psychoeducation or psychosocial intervention to improve sleep in short sleepers on reducing mortality and the incidence of these health outcomes should be examined in general community settings.

Authors' contributions

All authors contributed to the manuscript as follows:

OI designed the study, developed technical materials, acquired data, and interpreted the data. MJ designed the study, developed technical materials, acquired data, and interpreted the data. NW designed the study, interpreted the data, and drafted the manuscript.

YK obtained funding, conceived the study, designed the study, and interpreted the data.

All authors have revised the important intellectual content critically, have read and approved the final manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

Funding sources

This study was funded by Health Labor Sciences Research Grant (H25-JUNKANKITOU-IPPAN-007) from the Japanese Ministry of Health Labor and Welfare.

The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Conflict of interest

The authors have no conflicts of interests to declare that may be affected by the publication of the paper. Other conflicts of interests are as follows:

OI has research funds from the Japanese Ministry of Health Labor and Welfare and the Japanese Ministry of Education, Science, and Technology.

MJ has research funds from the Japanese Ministry of Health Labor.

NW has received research funds from the Japanese Ministry of Health Labor and Welfare, the Japanese Ministry of Education, Science, and Technology and National Center of Neurology and Psychiatry, Intramural Research Grant for Neurological and Psychiatric Disorders. He has also received royalties from Sogensha, Paquet, and Akatsuki.

YK has research funds from the Japanese Ministry of Health Labor and Welfare and the Japanese Ministry of Education, Science, and Technology. He has also received royalties from Dai-Nippon Sumitomo, MSD, Otsuka, Eisai, Pfizer, and Takeda during the last 5 years.

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2016.08.006.

Acknowledgments

This work was performed at Oita University, Nihon University, and Kyoto University, Japan.

Appendix A. Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.sleep.2016.08.006.

References

[1] Ford ES, Cunningham TJ, Croft JB. Trends in self-reported sleep duration among US adults from 1985 to 2012. Sleep 2015;38:829–32.
[2] Bin YS, Marshall NS, Glozier N. Sleeping at the limits: the changing prevalence of short and long sleep durations in 10 countries. Am J Epidemiol 2013;177:826–33.
[3] Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. J Sleep Res 2009;18:148–58.
[4] Cappuccio FP, D'Elia L, Strazzullo P, et al. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. Sleep 2010;33:585–92.
[5] Guo X, Zheng L, Wang J, et al. Epidemiological evidence for the link between sleep duration and high blood pressure: a systematic review and meta-analysis. Sleep Med 2013;14:324–32.
[6] Shen X, Wu Y, Zhang D. Nighttime sleep duration, 24-hour sleep duration and risk of all-cause mortality among adults: a meta-analysis of prospective cohort studies. Sci Rep 2016;6:21480.
[7] Wang Y, Mei H, Jiang YR, et al. Relationship between duration of sleep and hypertension in adults: a meta-analysis. J Clin Sleep Med 2015;11:1047–56.
[8] Cappuccio FP, Cooper D, Delia L, et al. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J 2011;32:1484–92.
[9] Leng Y, Cappuccio FP, Wainwright NW, et al. Sleep duration and risk of fatal and nonfatal stroke: a prospective study and meta-analysis. Neurology 2015;84:1072–9.
[10] Cappuccio FP, D'Elia L, Strazzullo P, et al. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010;33:414–20.
[11] Shan Z, Ma H, Xie M, et al. Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies. Diabetes Care 2015;38:529–37.
[12] Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. Sleep Med 2014;15:1456–62.
[13] Chaput JP, McNeil J, Despres JP, et al. Short sleep duration as a risk factor for the development of the metabolic syndrome in adults. Prev Med 2013;57:872–7.
[14] Sands MR, Lauderdale DS, Liu K, et al. Short sleep duration is associated with carotid intima-media thickness among men in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Stroke 2012;43:2858–64.
[15] King CR, Knutson KL, Rathouz PJ, et al. Short sleep duration and incident coronary artery calcification. JAMA 2008;300:2859–66.
[16] Hammond EC. Some preliminary findings on physical complaints from a prospective study of 1,064,004 men and women. Am J Public Health Nations Health 1964;54:11–23.
[17] Kaplan GA, Seeman TE, Cohen RD, et al. Mortality among the elderly in the Alameda County Study: behavioral and demographic risk factors. Am J Public Health 1987;77:307–12.
[18] Amagai Y, Ishikawa S, Gotoh T, et al. Sleep duration and mortality in Japan: the Jichi Medical School Cohort Study. J Epidemiol 2004;14:124–8.
[19] Tamakoshi A, Ohno Y, Group JS. Self-reported sleep duration as a predictor of all-cause mortality: results from the JACC study, Japan. Sleep 2004;27:51–4.
[20] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
[21] Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
[22] Zizi F, Pandey A, Murrray-Bachmann R, et al. Race/ethnicity, sleep duration, and diabetes mellitus: analysis of the National Health Interview Survey. Am J Med 2012;125:162–7.
[23] Adenekan B, Pandey A, McKenzie S, et al. Sleep in America: role of racial/ethnic differences. Sleep Med Rev 2013;17:255–62.
[24] Lauderdale DS, Knutson KL, Yan LL, et al. Objectively measured sleep characteristics among early-middle-aged adults: the CARDIA study. Am J Epidemiol 2006;164:5–16.
[25] Silva GE, Goodwin JL, Sherrill DL, et al. Relationship between reported and measured sleep times: the sleep heart health study (SHHS). J Clin Sleep Med 2007;3:622–30.
[26] Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008.
[27] Aziz O, Constantinides V, Tekkis PP, et al. Laparoscopic versus open surgery for rectal cancer: a meta-analysis. Ann Surg Oncol 2006;13:413–24.
[28] Millett GA, Flores SA, Marks G, et al. Circumcision status and risk of HIV and sexually transmitted infections among men who have sex with men: a meta-analysis. JAMA 2008;300:1674–84.
[29] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
[30] Aburto NJ, Hanson S, Gutierrez H, et al. Effect of increased potassium intake on cardiovascular risk factors and disease: systematic review and meta-analyses. BMJ 2013;346:f1378.
[31] Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–34.
[32] Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56:455–63.
[33] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78–84.
[34] Borenstein M, Hedges LV, Higgins JPT, et al. PART 4: meta-regression is observational. In: Borenstein M, Hedges LV, Higgins JPT, et al., editors. Regression in meta-analysis. Englewood (NJ): Biostat, Inc.; 2015. p. 19–20.
[35] Borenstein M, Hedges L, Higgins J, et al. Comprehensive meta analysis, Version 3. 3.3.070 ed. Englewood (NJ): Biostat, Inc.; 2014.
[36] Tsubono Y, Fukao A, Hisamichi S. Health practices and mortality in a rural Japanese population. Tohoku J Exp Med 1993;171:339–48.
[37] Kojima M, Wakai K, Kawamura T, et al. Sleep patterns and total mortality: a 12-year follow-up study in Japan. J Epidemiol 2000;10:87–93.
[38] Seki N. Relationships between walking hours, sleeping hours, meaningfulness of life (ikigai) and mortality in the elderly: prospective cohort study. Nihon Eiseigaku Zasshi 2001;56:535–40. in Japanese.
[39] Heslop P, Smith GD, Metcalfe C, et al. Sleep duration and mortality: the effect of short or long sleep duration on cardiovascular and all-cause mortality in working men and women. Sleep Med 2002;3:305–14.
[40] Mallon L, Broman JE, Hetta J. Sleep complaints predict coronary artery disease mortality in males: a 12-year follow-up study of a middle-aged Swedish population. J Intern Med 2002;251:207–16.
[41] Goto A, Yasumura S, Nishise Y, et al. Association of health behavior and social role with total mortality among Japanese elders in Okinawa, Japan. Aging Clin Exp Res 2003;15:443–50.
[42] Patel SR, Ayas NT, Malhotra MR, et al. A prospective study of sleep duration and mortality risk in women. Sleep 2004;27:440–4.
[43] Ferrie JE, Shipley MJ, Cappuccio FP, et al. A prospective study of change in sleep duration: associations with mortality in the Whitehall II cohort. Sleep 2007;30:1659–66.
[44] Lan TY, Lan TH, Wen CP, et al. Nighttime sleep, Chinese afternoon nap, and mortality in the elderly. Sleep 2007;30:1105–10.
[45] Gangwisch JE, Heymsfield SB, Boden-Albala B, et al. Sleep duration associated with mortality in elderly, but not middle-aged, adults in a large US sample. Sleep 2008;31:1087–96.
[46] Ikehara S, Iso H, Date C, et al. Association of sleep duration with mortality from cardiovascular disease and other causes for Japanese men and women: the JACC study. Sleep 2009;32:295–301.
[47] Mallon L, Broman JE, Hetta J. Is usage of hypnotics associated with mortality? Sleep Med 2009;10:279–86.
[48] Stone KL, Ewing SK, Ancoli-Israel S, et al. Self-reported sleep and nap habits and risk of mortality in a large cohort of older women. J Am Geriatr Soc 2009;57:604–11.
[49] Suzuki E, Yorifuji T, Ueshima K, et al. Sleep duration, sleep quality and cardiovascular disease mortality among the elderly: a population-based cohort study. Prev Med 2009;49:135–41.
[50] Chien KL, Chen PC, Hsu HC, et al. Habitual sleep duration and insomnia and the risk of cardiovascular events and all-cause death: report from a community-based cohort. Sleep 2010;33:177–84.
[51] Mesas AE, Lopez-Garcia E, Leon-Munoz LM, et al. Sleep duration and mortality according to health status in older adults. J Am Geriatr Soc 2010;58:1870–7.
[52] Vgontzas AN, Liao D, Pejovic S, et al. Insomnia with short sleep duration and mortality: the Penn State cohort. Sleep 2010;33:1159–64.
[53] Castro-Costa E, Dewey ME, Ferri CP, et al. Association between sleep duration and all-cause mortality in old age: 9-year follow-up of the Bambui Cohort Study, Brazil. J Sleep Res 2011;20:303–10.
[54] Kripke DF, Langer RD, Elliott JA, et al. Mortality related to actigraphic long and short sleep. Sleep Med 2011;12:28–33.
[55] Kronholm E, Laatikainen T, Peltonen M, et al. Self-reported sleep duration, all-cause mortality, cardiovascular mortality and morbidity in Finland. Sleep Med 2011;12:215–21.
[56] Cohen-Mansfield J, Perach R. Sleep duration, nap habits, and mortality in older persons. Sleep 2012;35:1003–9.
[57] Ensrud KE, Blackwell TL, Ancoli-Israel S, et al. Sleep disturbances and risk of frailty and mortality in older men. Sleep Med 2012;13:1217–25.
[58] Chen HC, Su TP, Chou P. A nine-year follow-up study of sleep patterns and mortality in community-dwelling older adults in Taiwan. Sleep 2013;36:1187–98.
[59] Garde AH, Hansen AM, Holtermann A, et al. Sleep duration and ischemic heart disease and all-cause mortality: prospective cohort study on effects of tranquilizers/hypnotics and perceived stress. Scand J Work Environ Health 2013;39:550–8.
[60] Hale L, Parente V, Dowd JB, et al. Fibrinogen may mediate the association between long sleep duration and coronary heart disease. J Sleep Res 2013;22:305–14.
[61] Kakizaki M, Kuriyama S, Nakaya N, et al. Long sleep duration and cause-specific mortality according to physical function and self-rated health: the Ohsaki Cohort Study. J Sleep Res 2013;22:209–16.

[62] Kim Y, Wilkens LR, Schembre SM, et al. Insufficient and excessive amounts of sleep increase the risk of premature death from cardiovascular and other diseases: the Multiethnic Cohort Study. Prev Med 2013;57:377–85.
[63] Li Y, Zhang X, Winkelman JW, et al. Association between insomnia symptoms and mortality: a prospective study of U.S. men. Circulation 2014;129:737–46.
[64] Magee CA, Holliday EG, Attia J, et al. Investigation of the relationship between sleep duration, all-cause mortality, and preexisting disease. Sleep Med 2013;14:591–6.
[65] Yeo Y, Ma SH, Park SK, et al. A prospective cohort study on the relationship of sleep duration with all-cause and disease-specific mortality in the Korean Multi-center Cancer Cohort study. J Prev Med Public Health 2013;46:271–81.
[66] Bellavia A, Akerstedt T, Bottai M, et al. Sleep duration and survival percentiles across categories of physical activity. Am J Epidemiol 2014;179:484–91.
[67] Rod NH, Kumari M, Lange T, et al. The joint effect of sleep duration and disturbed sleep on cause-specific mortality: results from the Whitehall II cohort study. PLoS ONE 2014;9. e91965.
[68] Xiao Q, Keadle SK, Hollenbeck AR, et al. Sleep duration and total and cause-specific mortality in a large US cohort: interrelationships with physical activity, sedentary behavior, and body mass index. Am J Epidemiol 2014;180:997–1006.
[69] Zawisza K, Tobiasz-Adamczyk B, Galas A, et al. Sleep duration and mortality among older adults in a 22-year follow-up study: an analysis of possible effect modifiers. Eur J Ageing 2015;12:119–29.
[70] Zuurbier LA, Luik AI, Hofman A, et al. Fragmentation and stability of circadian activity rhythms predict mortality: the Rotterdam study. Am J Epidemiol 2015;181:54–63.
[71] Hall MH, Smagula SF, Boudreau RM, et al. Association between sleep duration and mortality is mediated by markers of inflammation and health in older adults: the Health, Aging and Body Composition Study. Sleep 2015;38:189–95.
[72] Cai H, Shu XO, Xiang YB, et al. Sleep duration and mortality: a prospective study of 113,138 middle-aged and elderly Chinese men and women. Sleep 2015;38:529–36.
[73] Grandner MA, Hale L, Moore M, et al. Mortality associated with short sleep duration: the evidence, the possible mechanisms, and the future. Sleep Med Rev 2010;14:191–203.
[74] Benedict C, Hallschmid M, Lassen A, et al. Acute sleep deprivation reduces energy expenditure in healthy men. Am J Clin Nutr 2011;93:1229–36.
[75] Broussard JL, Ehrmann DA, Van Cauter E, et al. Impaired insulin signaling in human adipocytes after experimental sleep restriction: a randomized, cross-over study. Ann Intern Med 2012;157:549–57.
[76] Buxton OM, Cain SW, O'Connor SP, et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. Sci Transl Med 2012;4:129ra43.
[77] Taheri S, Lin L, Austin D, et al. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med 2004;1:e62.
[78] Spiegel K, Tasali E, Penev P, et al. Brief communication: sleep curtailment in healthy young men is associated with decreased leptin levels, elevated ghrelin levels, and increased hunger and appetite. Ann Intern Med 2004;141:846–50.
[79] Van Cauter E, Holmback U, Knutson K, et al. Impact of sleep and sleep loss on neuroendocrine and metabolic function. Horm Res 2007;67(Suppl. 1):2–9.
[80] Ogawa Y, Kanbayashi T, Saito Y, et al. Total sleep deprivation elevates blood pressure through arterial baroreflex resetting: a study with microneurographic technique. Sleep 2003;26:986–9.
[81] Gangwisch JE. Epidemiological evidence for the links between sleep, circadian rhythms and metabolism. Obes Rev 2009;10:37–45.
[82] Wu JQ, Appleman ER, Salazar RD, et al. Cognitive behavioral therapy for insomnia comorbid with psychiatric and medical conditions: a meta-analysis. JAMA Intern Med 2015;175:1461–72.
[83] Zachariae R, Lyby MS, Ritterband LM, et al. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia – a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2015;30:1–10.
[84] Grandner MA, Patel NP. From sleep duration to mortality: implications of meta-analysis and future directions. J Sleep Res 2009;18:145–7.
[85] Stranges S, Dorn JM, Shipley MJ, et al. Correlates of short and long sleep duration: a cross-cultural comparison between the United Kingdom and the United States: the Whitehall II Study and the Western New York Health Study. Am J Epidemiol 2008;168:1353–64.
[86] Krueger PM, Friedman EM. Sleep duration in the United States: a cross-sectional population-based study. Am J Epidemiol 2009;169:1052–63.
[87] Kurina LM, McClintock MK, Chen JH, et al. Sleep duration and all-cause mortality: a critical review of measurement and associations. Ann Epidemiol 2013;23:361–70.
[88] Grandner MA, Drummond SPA. Who are the long sleepers? Towards an understanding of the mortality relationship. Sleep Med Rev 2007;11:341–60.
[89] Liu TZ, Xu C, Rota M, et al. Sleep duration and risk of all-cause mortality: a flexible, non-linear, meta-regression of 40 prospective cohort studies. Sleep Med Rev 2016. http://dx.doi.org/10.1016/j.smrv.2016.02.005.
[90] Wang Q, Xi B, Liu M, et al. Short sleep duration is associated with hypertension risk among adults: a systematic review and meta-analysis. Hypertens Res 2012;35:1012–8.
[91] Araghi MH, Thomas GN, Taheri S. The potential impact of sleep duration on lipid biomarkers of cardiovascular disease. Clin Lipidol 2012;7:443–53.
[92] Alvarez GG, Ayas NT. The impact of daily sleep duration on health: a review of the literature. Prog Cardiovasc Nurs 2004;19:56–9.
[93] Parthasarathy S, Vasquez MM, Halonen M, et al. Persistent insomnia is associated with mortality risk. Am J Med 2015;128:268–75. e2.
[94] Ho FY, Chung KF, Yeung WF, et al. Self-help cognitive-behavioral therapy for insomnia: a meta-analysis of randomized controlled trials. Sleep Med Rev 2015;19:17–28.
[95] Petrov ME, Kim Y, Lauderdale D, et al. Longitudinal associations between objective sleep and lipids: the CARDIA study. Sleep 2013;36:1587–95.

BJPsych | The British Journal of Psychiatry (2017)
210, 190–196. doi: 10.1192/bjp.bp.116.187773

Review article

Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials

Toshi A. Furukawa,* Erica S. Weitz,* Shiro Tanaka,* Steven D. Hollon, Stefan G. Hofmann, Gerhard Andersson, Jos Twisk, Robert J. DeRubeis, Sona Dimidjian, Ulrich Hegerl, Roland Mergl, Robin B. Jarrett, Jeffrey R. Vittengl, Norio Watanabe and Pim Cuijpers

Background
The influence of baseline severity has been examined for antidepressant medications but has not been studied properly for cognitive–behavioural therapy (CBT) in comparison with pill placebo.

Aims
To synthesise evidence regarding the influence of initial severity on efficacy of CBT from all randomised controlled trials (RCTs) in which CBT, in face-to-face individual or group format, was compared with pill-placebo control in adults with major depression.

Method
A systematic review and an individual-participant data meta-analysis using mixed models that included trial effects as random effects. We used multiple imputation to handle missing data.

Results
We identified five RCTs, and we were given access to individual-level data ($n = 509$) for all five. The analyses revealed that the difference in changes in Hamilton Rating Scale for Depression between CBT and pill placebo was not influenced by baseline severity (interaction $P = 0.43$). Removing the non-significant interaction term from the model, the difference between CBT and pill placebo was a standardised mean difference of −0.22 (95% CI −0.42 to −0.02, $P = 0.03$, $I^2 = 0\%$).

Conclusions
Patients suffering from major depression can expect as much benefit from CBT across the wide range of baseline severity. This finding can help inform individualised treatment decisions by patients and their clinicians.

Declaration of interest
T.A.F. has received lecture fees from Eli Lilly, Meiji, Mochida, MSD, Otsuka, Pfizer and Tanabe-Mitsubishi and consultancy fees from Sekisui Chemicals and Takeda Science Foundation. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha publishers. He has received grant or research support from the Japanese Ministry of Education, Science and Technology, the Japanese Ministry of Health, Labour and Welfare, the Japan Society for the Promotion of Science, the Japan Foundation for Neuroscience and Mental Health, Mochida and Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. U.H. has received lecture fees from Bristol-Myers Squibb, Medice Arzneimittel, Novartis and Roche Pharma and was an advisory board member for Lilly, Lundbeck, Takeda Pharmaceuticals, Servier and Otsuka Pharma. R.M. has served as a consultant for Nycomed, a Takeda company. R.B.J. received support from the NIMH for the trial cited here (R0MH45043) and is a paid consultant to NIH including the NIMH and to UpToDate. J.R.V. is a paid consultant to UpToDate. N.W. has received royalties from Sogensha and Paquet, and speaker fees and research funds from Asahi Kasei, Dai-Nippon Sumitomo, Eli Lilly, GlaxoSmithKline, Janssen, Meiji, MSD, Otsuka and Pfizer.

Antidepressant medications and psychotherapies, especially cognitive–behavioural therapy (CBT), constitute two major options as empirically supported treatments for adults with major depression.[1,2] Psychotherapies have traditionally been believed to be better suited for treating people with mild depression. For example, the third edition of the practice guidelines for major depression published by the American Psychiatric Association states: 'Use of a depression-focused psychotherapy alone is recommended as an initial treatment choice for patients with mild to moderate major depressive disorder'. However, controversies have arisen recently with regard to the influence of baseline severity on the efficacy of psychotherapies for depression. One study-level analysis of trials comparing various psychotherapies, including CBT, interpersonal psychotherapy and supportive therapy, against pill placebo did not find an association between the study's effect size and mean baseline depression severity.[3] Yet another analysis at the study level found no effect of pre-treatment depression scores on the effect size of different psychotherapies over various control conditions (including waiting list, no treatment or pill placebo), but did find a clear indication that the effect size was bigger and statistically significant in high-severity patients and smaller and non-significant in low-severity patients in the small subset of studies that had examined such differences within each study.[4] However, these study-level analyses are limited in statistical power and are subject to the ecological fallacy that the relationships observed at the group level may not reflect the true relationships at the individual level.[5] Meta-analyses based on individual-participant data (IPD-MA) are necessary to examine patient-level effect moderators, including baseline severity. An IPD-MA of low-intensity psychotherapies (such as guided self-help bibliotherapy and internet-delivered therapy: 16 randomised controlled trails (RCTs), 2470 patients) found a significant albeit small interaction between baseline severity and treatment effect, suggesting that patients who are

*These authors contributed equally to the work.

initially more depressed demonstrate slightly larger treatment effects with such low-intensity interventions.[6]

Given the direct relevance of baseline severity with respect to tailoring treatment and informed consent for patients with mild, moderate and severe symptoms, the possibility of the influence of baseline severity on efficacy of treatments for adults with major depression merits further investigation. Such investigation needs to be methodologically rigorous in applying individual-participant-level meta-analysis and procedurally comprehensive in identifying and assembling individual data to be so analysed. The current study aims to systematically identify all RCTs examining CBT in a face-to-face format in comparison with pill placebo in the treatment of adults with major depression, retrieve their individual-level data and apply IPD-MA to examine the possible influence of baseline depression severity on its efficacy. CBT is by far the best-researched and widely practised form of psychotherapy for depression.[7,8] We chose pill placebo as the control condition in this IPD-MA mainly for two reasons. First, traditional control conditions used in psychotherapy trials, such as waiting list, no treatment or treatment as usual, are too heterogeneous and may be affected by publication bias.[9,10] Second, and more importantly, the pill-placebo condition will control for non-specific placebo effects including expectation, attention and support and will offer a common comparison condition for both antidepressants and psychotherapies. Our hypothesis was that increases in baseline severity would be associated with increases in the advantages of CBT over pill placebo among adult patients with major depression.

Method

Selection of studies and data extraction

The eligibility criteria for the present IPD-MA were as follows:

(a) RCT;

(b) the participants were adult patients with major depressive disorder, diagnosed according to any operational diagnostic criteria including the Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-5 or ICD-10;

(c) the intervention was face-to-face CBT in individual or group format. We defined CBT as a psychotherapy focused on the impact of a patient's dysfunctional thoughts on his/her current behaviour and future functioning and aimed at evaluating, challenging and modifying such dysfunctional beliefs (cognitive restructuring). We included two subtypes thereof, namely one in which cognitive restructuring is the core element and one in which it is an important component but in which at least two other components such as behavioural activation, social skills training, relaxation or coping skills also have a prominent role;[11]

(d) the non-specific control was pill placebo. The trial could have other active or non-active intervention arms.

We conducted a systematic and cumulative search of all RCTs of psychotherapies for depression (www.evidencebasedpsychotherapies.org). The searches included PubMed, PsycInfo, EMBASE and the Cochrane Central Register of Controlled Trials using keywords indicative of psychological treatment and depression, and were supplemented by examination of published meta-analyses of psychological treatments for depression, up to January 2014. An updated search was conducted on 16 January 2015 with PubMed and the Cochrane Central Register. We identified the relevant studies satisfying our eligibility criteria from this database and then asked the primary authors of all the identified studies to contribute individual-participant data for our primary and secondary outcomes (see below) as well as any possible effect moderators that each study may have measured. Study quality was assessed using four criteria from the Cochrane Collaboration's risk of bias tool for random sequence generation, concealment of random allocation, masking of assessors and complete outcome reporting. When the information provided in the publications was not clear, we sought information from the original study authors.

Outcomes

Our pre-specified primary outcome measure was change in the Hamilton Rating Scale for Depression (HRSD)[12] scores from baseline to the last study time point at which valid comparison between CBT and pill placebo could be made. Two studies employed the pill-placebo arm only through week 8 of a 16-week study,[13,14] in which case we were able to compare CBT against pill placebo up to that time point only. Our secondary outcome was change in the Beck Depression Inventory (BDI).[15]

Analyses

We first verified the integrity of the individual-participant data by comparing their descriptive statistics against the published figures. We conducted IPD-MA to estimate the difference in symptom change scores between the CBT and pill-placebo arms and to examine the influence of the baseline symptom severity on this difference. Because the original studies used different versions of the HRSD and BDI with different item questions and with different numbers of items, such continuous outcomes were standardised within each study, that is, the individual patients' raw scores at baseline and at end-point were divided by the study-specific standard deviations of end-point scores. After this standardisation of scores from different scales, when we graphically presented the relationship between the scores at baseline and the changes in scores, we subtracted the grand mean of all the studies' standardised scores from individual standardised scores in order to anchor the horizontal and vertical axes in the figure. The relationships between these standardised baseline symptom severity scores and the differences in symptom change scores were examined as the interaction term between baseline severity and treatment in the mixed models. When this interaction term was not statistically significant, we dropped the interaction term and estimated the main effect of treatment. The mixed models included trial effects as random effects throughout.

Missing data in HRSD and BDI scores were handled by the multiple imputation method using study, treatment, age, gender, minority status, marital status, employment, education, age at onset, diagnosis of depression subtype, presence of previous episodes, length of current episode, comorbid anxiety disorders, past substance use disorders, length of treatment and HRSD or BDI scores at baseline and at the last study time point as predictors. Multiple imputation is one of the currently recommended alternatives to the long-used last-observation-carried-forward (LOCF) method.[16] It has some advantages over hierarchical linear modelling as well as LOCF because auxiliary variables that are not included in the final analysis model can be used in the imputation model.[17]

We also conducted sensitivity analyses (a) excluding studies that continued the pill-placebo arm only through the mid-point of acute treatment, (b) excluding studies that administered CBT in a group format, (c) including patients diagnosed with dysthymia or minor depression at baseline, (d) employing the multiple imputation using the predictors other than treatment, and (e) restricted to only those participants who completed the treatment.

All reported *P*-values are two-tailed and $P<0.05$ was chosen as the threshold for statistical significance. An academic statistician (S.T.) conducted all statistical analyses using SAS version 9.3. This study was conducted in accordance with PRISMA and PRISMA-IPD guidelines.[18]

Results

Studies and participants

We found five RCTs that satisfied our eligibility criteria (online Fig. DS1).[13,14,19–21] One study[21] included some patients with minor depression or dysthymia: the latter patients were excluded in our primary analyses but were included in a sensitivity analysis to give a wider range of baseline severity. All the primary authors and their collaborators agreed to contribute to the present study by providing individual-level data for the primary and secondary outcomes as well as possible effect moderators that they had measured. The data from the NIMH Treatment of Depression Collaborative Program[19] were in the public domain and available for inclusion.

Table 1 lists the characteristics of the five RCTs that compared CBT against pill placebo. We obtained individual-patient data for 509 patients with major depression and 46 patients with minor depression and dysthymia. The mean age of the patients with major depression participating in these trials was in the 30s or 40s, with a preponderance of women. In addition to CBT and pill-placebo arms, the comparison of interest in the present study, all the studies had an antidepressant medication arm and three studies also included a different psychotherapy from CBT. One study[21] used the original 17-item version of HRSD, one study[20] the 21-item version, and the remaining three studies[13,14,19] employed the modified HRSD with additional items for atypical symptoms. Two studies[19,20] used the original BDI and three studies[13,14,21] used the second version.

All the studies were rated at low risk of bias for random sequence generation, allocation concealment, masking and incomplete outcome reporting, except for one study whose risk of bias was rated unclear as to allocation concealment but low for all the other aspects.[14] Raters masked to treatment assignment administered the HRSD in all studies. As a self-report measure, patients knew whether they were receiving CBT or were in a 'medication' condition when they completed the BDI; however, they were masked as to whether they were taking an active medication or a pill placebo.

Baseline severity and symptom change in CBT and pill placebo

Figure 1 depicts observed and estimated changes in standardised HRSD scores in the CBT or pill-placebo arms. The end-point standard deviations used to standardise the HRSD score were 5.8 for DeRubeis *et al*,[13] 6.6 for Dimidjian *et al*,[14] 6.3 for Elkin *et al*,[19] 6.2 for Hegerl *et al*[21] and 7.3 for Jarrett *et al*,[20] and we used the grand means of the standardised scores from all the studies, 3.2 for baseline scores and 1.3 for change scores, as subtrahends in scaling the axes in Fig. 1. The individual participant-level meta-analysis revealed that the differences in changes in HRSD between CBT and pill-placebo arms were not influenced by baseline HRSD severity. The slopes in the CBT and placebo arms were estimated to be −0.615 (95% C1 −0.839 to −0.392, $P<0.01$) and −0.526 (95% CI −0.733 to −0.319, $P<0.01$), respectively and the interaction term, i.e. the difference between the slopes, was −0.089 (−0.311 to 0.133, $P=0.43$).

The interaction between baseline severity and treatment condition was also not significant with change in BDI as the dependent variable. The slopes in the CBT and pill-placebo arms were estimated to be −0.541 (95% CI −0.724 to −0.358, $P<0.01$) and −0.615 (95% CI −0.793 to −0.438, $P<0.01$), respectively and the interaction term was 0.074 (−0.175 to 0.323, $P=0.56$). Removing the higher-order non-significant interaction terms from the models, the IPD-MA revealed that the standardised mean difference (SMD) between CBT and pill-placebo arms was −0.220 (95% CI −0.419 to −0.022, $P=0.03$, $I^2=0\%$) for the HRSD and −0.046 (95% CI −0.264 to 0.172, $P=0.68$, $I^2=0\%$) for the BDI (Table 2).

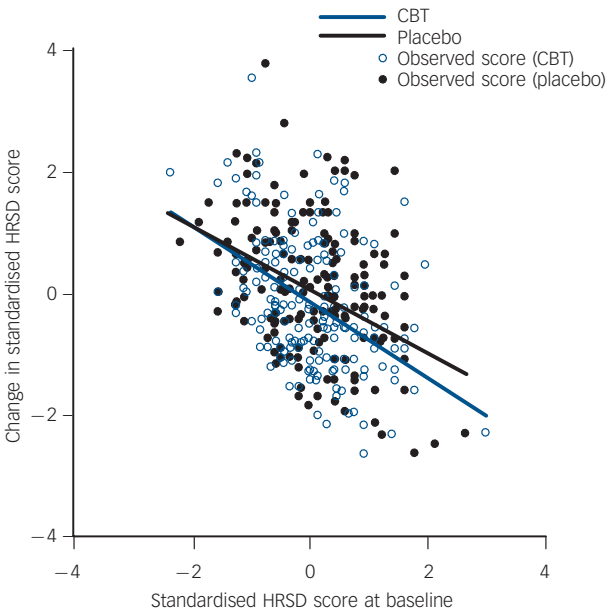


Fig. 1 Observed (dots) and estimated (lines) changes in standardised scores on Hamilton Rating Scale for Depression (HRSD) in the cognitive–behavioural therapy (CBT) and pill-placebo arms.

Sensitivity analyses

We also conducted sensitivity analyses to examine the robustness of our primary analyses by (a) excluding DeRubeis *et al*[13] and Dimidjian *et al*'s[14] studies that had the pill-placebo arm only up to the mid-point of the study, (b) excluding Hegerl *et al*'s study[21] that administered group-format CBT, (c) including patients diagnosed with dysthymia or minor depression at baseline, (d) multiple imputation using the predictors other than treatment and (e) restricting the sample to participants who completed the treatment only (Table 3). The interaction terms were all statistically non-significant. The treatment effects in models in which the interaction term was excluded were statistically significant in the sensitivity analyses (c) and (e) above but not in (a), (b) or (d), likely because of reduced sample sizes or to residual bias related to missing data. However, the estimated effects ranged between −0.16 and −0.27, and in all instances their 95% confidence intervals overlapped with those from the primary analysis (−0.22, 95% CI −0.42 to −0.02).

Discussion

Main findings

We identified five pill-placebo-controlled randomised studies of CBT conducted to date, obtained their individual-level data (509 patients with major depression), and conducted an IPD-MA to examine the influence of baseline depression severity on

Table 1 Characteristics of the included five randomised controlled trials (RCTs) and their participants included in the current meta-analysis

Study	Key eligibility criteria for the original study	Randomisation arms in the original studies	Characteristics of the participants included in the present IPD-MA	Effect size of CBT over pill placebo according to the original publications and used in a previous aggregate data meta-analysis[27]
DeRubeis *et al*[13] (2005)	Major depression with baseline modified HRSD-17[a] score ⩾20	CBT ($n=60$, 16 weeks: 50 min sessions twice weekly for the first 4 weeks, once or twice weekly for the next 8 weeks, and once weekly for the final 4 weeks) Paroxetine ($n=120$, 16 weeks: once weekly for the first 4 weeks, biweekly thereafter; initial session lasted 30–45 min, and 20 min thereafter) Pill placebo ($n=60$, 8 weeks: same as paroxetine)	Diagnosis: major depression ($n=120$) Mean age: 41.0 Women, %: 59.2% Baseline modified HRSD-17[a] mean: 24.1 (s.d. = 3.4, range 20–36)	−0.44 ($P=0.09$) from HLM
Dimidjian *et al*[14] (2006)	Major depression with baseline modified HRSD-17[a] score ⩾14	CBT ($n=45$, 16 weeks: maximum of 24 50 min sessions, with sessions generally held twice weekly for the first 8 weeks and once weekly for the next 8 weeks) Behavioural activation ($n=43$, 16 weeks: same as CBT) Paroxetine ($n=100$, 16 weeks: once weekly for the first 4 weeks, biweekly thereafter; initial session lasted 30–45 min, and 30 min thereafter) Pill placebo ($n=53$, 8 weeks: same as paroxetine)	Diagnosis: major depression ($n=98$) Mean age: 39.8 Women, %: 72.5% Baseline modified HRSD-17[a] mean: 20.6 (s.d. = 4.5, range 14–34)	−0.40 (ns) from completers' data
Elkin *et al*[19] (1989)	Major depression with baseline modified HRSD-17[a] score of ⩾14	CBT ($n=59$, 16 weeks: 16 to 20 50 min sessions) IPT ($n=61$, 16 weeks: same as CBT) Imipramine ($n=57$, 16 weeks: 16–20 sessions; the initial session lasted 45–60 min and 20–30 min thereafter) Pill placebo ($n=62$, 16 weeks: same as imipramine)	Diagnosis: major depression ($n=121$) Mean age: 35.0 Women, %: 71.9% Baseline modified HRSD-17[a] mean: 20.7 (s.d. = 4.0, range 14–32)	−0.32 (ns) from LOCF data
Hegerl *et al*[21] (2010)	Minor depression, dysthymia or major depression with mild to moderate severity with baseline HRSD-17 score of 8–22	CBT ($n=61$, 10 weeks: 9 weekly 90 min group sessions) Sertraline ($n=83$, 10 weeks: 6 sessions) Pill placebo ($n=83$, 10 weeks: same as sertraline) Guided self-help ($n=59$, 10 weeks: same as CBT) Patients' choice (patient could choose either CBT or sertraline) ($n=82$)	Diagnosis: major depression ($n=98$), minor depression or dysthymia ($n=46$) Mean age: 47.9 Women, %: 70.4% Baseline HRSD-17 mean: 16.2 (s.d. = 4.4, range 5–27) for major depression; mean: 15.5 (s.d. = 4.6, range 3–23) for minor depression or dysthymia	−0.34 ($P=0.31$) from completers' data[b]
Jarrett *et al*[20] (1999)	Major depression with atypical features with baseline HRSD-21 score of ⩾14	CBT ($n=36$, 10 weeks: 20 twice-weekly sessions) Phenelzine ($n=36$, 10 weeks: 11 sessions) Pill placebo ($n=36$, 10 weeks: same as phenelzine)	Diagnosis: major depression ($n=72$) Mean age: 40.1 Women, %: 66.7% Baseline HRSD-21 mean: 17.9 (s.d. = 3.3, range 14–28)	−0.53 ($P<0.01$) from LOCF data

IPD-MA, individual-participant data meta-analysis; CBT, cognitive–behavioural therapy; HRSD, Hamilton Rating Scale for Depression; HLM, hierarchical linear modelling; IPT, interpersonal psychotherapy; ns, non-significant; LOCF, last-observation-carried-forward.
a. Modified by including items for hypersomnia, hyperphagia and weight gain.
b. Lynch *et al*[27] did not include this study, which was published after their systematic review. We included this reported effect size here for comparison.

Table 2 Treatment effects of cognitive–behavioural therapy (CBT) over pill placebo on different outcome measures

	P for interaction	*P* for main effect excluding the interaction term	Standardised mean difference (main effect)[a] (95% CI)
Hamilton Rating Scale for Depression change	0.43	0.03	−0.220 (−0.419 to −0.022)
Beck Depression Inventory change	0.56	0.68	−0.046 (−0.264 to 0.172)

a. Estimated by random-effects models after multiple imputation without the interaction term.

Table 3 Sensitivity analyses

	P for interaction	*P* for main effect excluding the interaction term	Standardised mean difference (main effect)[a] (95% CI)
Excluding two trials that continued the pill-placebo arm only through the mid-point of acute treatment and in which the comparison between CBT and placebo was therefore possible only up to the middle of the study[13,14]	0.21	0.12	−0.217 (−0.494 to 0.060)
Excluding a trial that administered CBT in group format[21] and limiting to CBT in individual format	0.67	0.08	−0.191 (−0.407 to 0.024)
Including minor depression and dysthymia	0.43	0.02	−0.230 (−0.431 to −0.030)
Multiple imputation using predictors other than treatment	0.34	0.13	−0.158 (−0.364 to 0.048)
Complete-case analysis	0.39	0.01	−0.266 (−0.480 to −0.053)

CBT, cognitive–behavioural therapy.
a. Estimated by random-effects models without the interaction term after multiple imputation except for the complete-case analysis.

its efficacy. There was no statistically significant influence of baseline depression severity, measured either with HRSD or BDI, on subsequent differential symptom change between CBT and pill placebo. The findings, thus, did not support our initial hypothesis that the efficacy of CBT over pill placebo would increase as baseline depression severity increased, when it was measured either by patient report or clinician ratings.

Finding that baseline severity has little influence on the efficacy of CBT can be said to be largely in accord with the previous literature, using study-level data[4,22] or individual-participant data.[6] Although a handful of studies that examined subgroup differences between low- *v.* high-severity groups have found evidence for moderation (with differences relative to pill placebo emerging only among patients with more severe depression) that was only the case for more purely behavioural interventions or interpersonal psychotherapy.[4] The IPD-MA that did find moderation by baseline severity concluded that the magnitude of interaction was small and that patients who were less severely depressed could derive as much benefit from low-intensity psychotherapies as did individuals who were more severely depressed.[6] Taken together, the evidence to date regarding the comparison of CBT *v.* pill placebo suggests that we can expect as much benefit from CBT across a wide range of baseline depression severity.

Excluding the non-significant interaction term and compared with pill placebo, CBT led to greater symptom reduction, on average, by an SMD of −0.22 (95% CI −0.42 to −0.02, $P=0.03$) on HRSD. The SMD in terms of BDI was −0.05 and was not statistically significant (95% CI −0.26 to 0.17, $P=0.68$). The HRSD was the prespecified primary outcome in this IPD-MA, as it was in all the included studies.[13,14,19–21] Generally, observer-rating scales including HRSD are said to be more sensitive to change than self-rating scales such as BDI.[23,24] Specifically in all the included studies, the HRSD was rated by assessors masked to all the treatments whereas the BDI was self-rated by participants who were masked to the distinctions in medications (active *v.* placebo) but not to those between CBT *v.* medications. It must also be pointed out that BDI was more often missing than HRSD in the original data-set (19.5% *v.* 13.7%).

Comparison with findings from other studies

The SMD of 0.22 found in the primary analysis of this meta-analysis is smaller than the previously estimated SMD of CBT over waiting-list conditions, the most common control condition in the psychotherapy trials, such as 0.82[25] or 0.85.[8] However, it must be remembered that the magnitude of effect is dependent not only on the effects of treatment but also on the effects of the control condition against which the therapy is compared. Waiting-list control conditions are known to lead to greater effect estimates than other control conditions, including clinical management that is accompanied by the provision of medication placebos.[10,23] In the control arms in four of the five trials in our data-set[13,14,19,20] the pharmacotherapists explicitly followed the clinical management protocol originally developed for the NIMH Treatment of Depression Collaborative Program[24] and in the remaining study[21] the psychiatrists' practices were in accord with this protocol, so that these pill-placebo conditions have included strong supportive components. Furthermore, the effect size estimates obtained in previous comparisons of psychotherapies with various control conditions may have been inflated by the inclusion of poor-quality studies[26] and by publication bias.[9] Our choice of the pill-placebo condition, which controls for the non-specific effects of expectation, attention and support, our comprehensive and systematic search of the relevant literature, and the high quality of the included studies including masked assessment of the outcome should have mitigated these problems.

Second, our estimate of the CBT–pill-placebo difference was also smaller than that reported in a previous study-level meta-analysis of pill-placebo-controlled trials. Lynch *et al*[27] reported an SMD of 0.41 (95% CI 0.21–0.61) between CBT and pill placebo from a study-level aggregate data meta-analysis of four of the five studies included in our IPD-MA.[13,14,19,20] This meta-analysis took effect size estimates for the included studies according to the original publications (see Table 1). One of the estimates was based on hierarchical linear modelling (HLM),[13] two others on the LOCF data[19,20] (although results from random regression were available from one[20] but were not used), and another on completers' data.[14] In the arms of the included studies, 15–64% of the randomised patients dropped out per arm and therefore

did not provide data at the end-point assessment. How drop-out was handled in the respective studies would be expected to influence estimates of the CBT *v.* pill-placebo difference. Although in the original publications, the trial authors may have used statistical approaches, including HLM[14,21] that are currently considered appropriate or were deemed to be so at the time they were published, the systematic review authors were obliged to extract data in the format amenable to meta-analysis, as detailed above. However, as methods have progressed today, neither the LOCF nor the completers' data would be considered the standard method to summarise and analyse data with missing values.[16] By contrast, we were able to meta-analyse the five studies with individual-participant data, which allowed us to use multiple imputation and test for effects using a mixed model. It is increasingly recognised that IPD-MA has many advantages, albeit with its own difficulties,[28] and some authors consider it to be the gold standard in evidence synthesis.[29]

Many questions remain when we try to factor antidepressant pharmacotherapy into the comparison between CBT and pill placebo. First, our estimated standardised effect size of 0.22 appears smaller than the publication-bias-free estimate of 0.31 of newer antidepressants over pill placebo.[30] However, it is problematic to compare these estimates directly with each other, as they were obtained from different original investigations. Moreover, our estimate for CBT *v.* pill-placebo difference is based on IPD-MA using multiple imputation, whereas that of the antidepressant *v.* pill-placebo difference derives from an aggregate data meta-analysis from published and unpublished trials, many of which used LOCF. Second, when directly compared in a study-level meta-analysis, the difference between CBT and second-generation antidepressants was not statistically significant.[31] A recent study applying IPD-MA to 1466 patients from 14 out of 24 identified relevant studies comparing CBT and pharmacotherapies revealed a small difference in effect size (SMD = 0.11, $P = 0.03$) but no significant differences in response or remission rates (odds ratio (OR) = 1.24, $P = 0.12$ and OR = 1.18, $P = 0.22$, respectively).[32] Lastly, the independence of the effect of CBT for depression from the baseline severity, as shown in the present analysis, is also in contrast to some claims to the contrary with regard to the effect of antidepressant drugs.[33] There is collateral evidence that the effect of antipsychotic drugs is also moderated by baseline psychotic severity.[34] However, the question of the influence of baseline symptom severity on antidepressant medications' efficacy is hardly closed,[35] and hence neither is the issue if such influences may differ between antidepressant medication and psychotherapies.

Limitations

Our study has limitations. First, detection of an interaction effect requires a large sample size. The present findings require replication with a larger sample of trials designed to test the interaction. The trials from which the data were extracted were not designed to test this hypothesis, but do provide a 'first look' into this question using all available data. Given the wide confidence interval around zero for the interaction coefficient in the present study, it is possible that our finding represents a type II error. Although none of the sensitivity analyses were suggestive of possible interaction effects, future studies may nonetheless still reveal an interaction that we failed to detect in the data available to us. It is disappointing that there were only 5 studies that compared CBT with pill placebo, whereas more than 20 studies compared it with antidepressant medications[32] and hundreds of studies have compared antidepressants with pill placebo.[36] Second, we were not able to examine important potential moderators such as therapist competence or other patient characteristics including comorbid personality disorder because such data were not consistently available in the obtained data-set although such covariates were requested in the source data from parent trials. Third, as the included studies used different versions of HRSD or BDI, we were obliged to standardise these severity measures using their means and standard deviations. This was unavoidable in order to synthesise different metrics across studies. Lastly, in combining data, we also ignored some design features from the parent studies that may or may not prove to be important in future understanding of antidepressant treatment effects (for example exclusive focus on subtypes of major depression in some studies, such as atypical depression,[20] primary care settings;[21] inclusion of a 'non-specific treatment run-in' prior to randomisation;[19,20] differential lengths of acute phase treatment; or inclusion of active psychotherapy arms other than CBT in the study). In view of today's growing emphasis on the importance of the totality of evidence[37] and hence systematic reviews including IPD-MAs,[29] researchers may benefit by making concerted efforts to measure core variables in their original research and then by contributing to the common database.[38]

Implications

Clinical implications of the present study may be as follows. First of all, patients and their clinicians can expect as much benefit from CBT for major depression across its wide range of baseline severity from mild through severe. Second, the benefit that can be expected was an effect size of 0.22, from which it can be calculated that the number needed to treat (NNT) is 12 for typical cases of major depression, for whom the expected placebo response rates may be between 30 and 50%.[39] This would compare favourably with the NNT of nine that can be expected for antidepressants with an effect size of 0.31 over placebo.[30] The difference in NNTs to obtain one more response with CBT or antidepressant medications than with pill placebo is therefore small enough to allow value judgements and preferences of individual patients to play a major role in treatment decision-making.

There are several research implications. We have been fortunate in having been able to identify and assemble all the relevant randomised evidence on the topic. We need to foster further collaborations in the field of psychotherapy research. Similar collaboration is burgeoning and is welcome in all fields of medicine.[38] It is also expected that such collaboration after completion of individual studies will increase the research community's awareness of the need for greater collaboration before individual studies by agreeing on common core variables to be assessed in the studies.

Toshi A. Furukawa, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Erica S. Weitz**, MA, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands; **Shiro Tanaka**, PhD, Department of Pharmacoepidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Steven D. Hollon**, PhD, Department of Psychology, Vanderbilt University, Nashville, Tennessee, USA; **Stefan G. Hofmann**, PhD, Department of Psychological and Brain Science, Boston University, Massachusetts, USA; **Gerhard Andersson**, PhD, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden and Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stokholm, Sweden; **Jos Twisk**, PhD, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands; **Robert J. DeRubeis**, PhD, Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; **Sona Dimidjian**, PhD, Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, USA; **Ulrich Hegerl**, MD, PhD, **Roland Mergl**, PhD, Department of Psychiatry and Psychotherapy, University of Leipzig, Leipzig, Germany; **Robin B. Jarrett**, PhD, Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, Texas, USA; **Jeffrey R. Vittengl**, PhD, Department of Psychology, Truman State University, Kirksville, Missouri, USA; **Norio Watanabe**, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Pim Cuijpers**, PhD, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands

Correspondence: Toshi A. Furukawa, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health. Yoshida Konoe-cho, Sakyo-ku, Kyoto, 606-8501 Japan. Email: furukawa@kuhp.kyoto-u.ac.jp

First received 13 May 2016, final revision 6 Sep 2016, accepted 10 Sep 2016

Funding

This study was supported in part by Grant-in-Aid for Challenging Exploratory Research from Japan Society for the Promotion of Science (JSPS) to T.A.F. and S.T. (No. 26670314) and by Health and Labour Sciences Research Grant to T.A.F. (H25-Seishin-Ippan-002). S.D.H. is supported by NIMH grant MH60713 and MH01697. S.G.H. receives support from NIH/NCCIH (R01AT007257), NIH/NIMH (R01MH099021, R34MH099311, R34MH086668, R21MH102646, R21MH101567, K23MH100259) and the Department of the Army for work unrelated to the studies reported in this article. R.J.D. is supported by NIMH grant MH60998. N.W. has research funds from the Japanese Ministry of Health Labor and Welfare and the Japanese Ministry of Education, Science and Technology. The funders had no role in the study design, the collection, analysis and interpretation of data, the writing of the report, or the decision to submit the article for publication.

References

1 National Institute for Health and Care Excellence. *Depression: The Treatment and Management of Depression in Adults (partial update of NICE clinical guideline 23)*. NICE, 2009.

2 American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (third edition). *Am J Psychiatry* 2010; **167** (suppl): 1–152.

3 Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol Med* 2013: 1–11.

4 Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J Consult Clin Psychol* 2010; **78**: 668–80.

5 Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; **26**: 2982–99.

6 Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ* 2013; **346**: f540.

7 Chen P, Furukawa TA, Shinohara K, Honyashiki M, Imai H, Ichikawa K, et al. Quantity and quality of psychotherapy trials for depression in the past five decades. *J Affect Disord* 2014; **165**: 190–5.

8 Barth J, Munder T, Gerger H, Nuesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med* 2013; **10**: e1001454.

9 Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive–behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry* 2010; **196**: 173–8.

10 Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand* 2014; **130**: 181–92.

11 Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol* 2008; **76**: 909–22.

12 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.

13 DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry* 2005; **62**: 409–16.

14 Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol* 2006; **74**: 658–70.

15 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psych* 1961; **4**: 561–71.

16 Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; **367**: 1355–60.

17 National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press, 2010.

18 Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015; **313**: 1657–65.

19 Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Arch Gen Psychiatry* 1989; **46**: 971–82.

20 Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Arch Gen Psychiatry* 1999; **56**: 431–7.

21 Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schutze M, Scheunemann W, et al. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *Int J Neuropsychopharmacol* 2010; **13**: 31–44.

22 Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol Med* 2014; **44**: 685–95.

23 Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol* 2013; **13**: 150.

24 Fawcett J, Epstein P, Fiester SJ, Elkin I, Autry JH. Clinical management–imipramine/placebo administration manual. NIMH Treatment of Depression Collaborative Research Program. *Psychopharmacol Bull* 1987; **23**: 309–24.

25 Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006; **26**: 17–31.

26 Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2010; **40**: 211–23.

27 Lynch D, Laws KR, McKenna PJ. Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. *Psychol Med* 2010; **40**: 9–24.

28 Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; **25**: 76–97.

29 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340**: c221.

30 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**: 252–60.

31 Amick HR, Gartlehner G, Gaynes BN, Forneris C, Asher GN, Morgan LC, et al. Comparative benefits and harms of second generation antidepressants and cognitive behavioral therapies in initial treatment of major depressive disorder: systematic review and meta-analysis. *BMJ* 2015; **351**: h6019.

32 Weitz ES, Hollon SD, Twisk J, van Straten A, Huibers MJ, David D, et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry* 2015; **72**: 1102–9.

33 Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**: 47–53.

34 Furukawa TA, Levine SZ, Tanaka S, Goldberg Y, Samara M, Davis JM, et al. Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. *JAMA Psychiatry* 2015; **72**: 14–21.

35 Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012; **69**: 572–9.

36 Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, Van Noord M, et al. Comparative benfits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med* 2011; **155**: 772–85.

37 Ioannidis JP, Karassa FB. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. *BMJ* 2010; **341**: c4875.

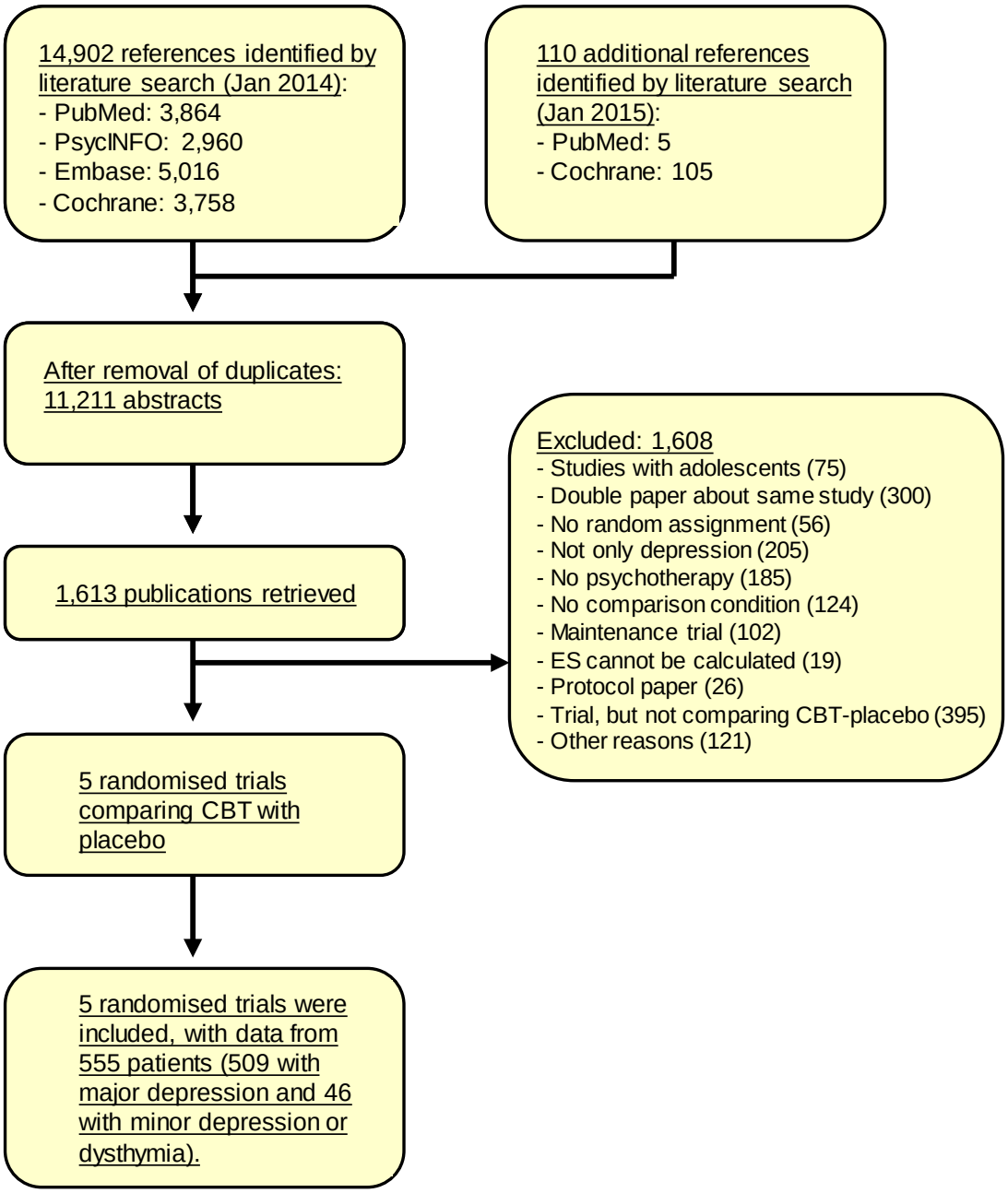
38 Institute of Medicine Committee on Strategies for Responsible Sharing of Clinical Trial Data. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*. National Academic Press, 2015.

39 Furukawa TA. From effect size into number needed to treat. *Lancet* 1999; **353**: 1680.

EXTRA CONTENT ONLINE

Data supplement to Furukawa et al. Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. British Journal of Psychiatry. doi: 10.1192/bjp.bp.116.187773

Fig. DS1 Study selection



Cochrane Library

Cochrane Database of Systematic Reviews

Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis (Protocol)

De Crescenzo F, Foti F, Ciabattini M, Del Giovane C, Watanabe N, Sañé Schepisi M, Quested DJ, Cipriani A, Barbui C, Amato L

De Crescenzo F, Foti F, Ciabattini M, Del Giovane C, Watanabe N, Sañé Schepisi M, Quested DJ, Cipriani A, Barbui C, Amato L.
Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis.
Cochrane Database of Systematic Reviews 2016, Issue 9. Art. No.: CD012364.
DOI: 10.1002/14651858.CD012364.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Protocol]

Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis

Franco De Crescenzo[1], Francesca Foti[2,3], Marco Ciabattini[4], Cinzia Del Giovane[5], Norio Watanabe[6], Monica Sañé Schepisi[7], Digby J Quested[8], Andrea Cipriani[9], Corrado Barbui[10], Laura Amato[11]

[1]Institute of Psychiatry and Psychology, Catholic University of the Sacred Heart, Rome, Italy. [2]Department of Psychology, "Sapienza" University of Rome, Rome, Italy. [3]IRCCS Santa Lucia Foundation, Rome, Italy. [4]Public Health, Tor Vergata University, Rome, Italy. [5]Italian Cochrane Centre, Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy. [6]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan. [7]Clinical Epidemiology, National Institute for Infectious Diseases, IRCCS L. Spallanzani, Rome, 00149, Italy. [8]Department of Psychiatry and Mental Health, Oxford Health NHS Trust, Oxford, UK. [9]Department of Psychiatry, University of Oxford, Oxford, UK. [10]Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy. [11]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy

Contact address: Franco De Crescenzo, Institute of Psychiatry and Psychology, Catholic University of the Sacred Heart, L.go A. Gemelli 8, Rome, 00168, Italy. decrescenzo.franco@gmail.com.

Editorial group: Cochrane Common Mental Disorders Group.
Publication status and date: Edited (no change to conclusions), published in Issue 9, 2016.

Citation: De Crescenzo F, Foti F, Ciabattini M, Del Giovane C, Watanabe N, Sañé Schepisi M, Quested DJ, Cipriani A, Barbui C, Amato L. Comparative efficacy and acceptability of pharmacological treatments for insomnia in adults: a systematic review and network meta-analysis. *Cochrane Database of Systematic Reviews* 2016, Issue 9. Art. No.: CD012364. DOI: 10.1002/14651858.CD012364.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

1) To compare individual pharmacological treatments for insomnia in adults in terms of:

- efficacy, measured as self-rated quality of sleep or satisfaction with sleep; and
- acceptability of treatment.

2) To generate a clinically-useful hierarchy of available pharmacological treatments for insomnia in adults, according to their efficacy and acceptability.

BACKGROUND

Description of the condition

Insomnia is characterised by difficulty in initiating and maintaining sleep, by early morning awakenings and by non-restorative

sleep, which leads to a condition of daytime blunting. Sleep disorders can lead to a reduction in cognitive, social-emotional, or occupational functioning.

Typical symptoms of insomnia related to sleep (i.e. difficulty initiating sleep, difficulty maintaining sleep, early morning awakenings, non-restorative sleep) are widespread, with an average point prevalence affecting about a third of the population. The prevalence tends to fall to 10% to 15% of the population when considering daytime impairment, and the prevalence falls further to a range of 6% to 10% if the main sets of diagnostic criteria are considered (Ohayon 2002; Morin 2006; Ohayon 2009). Daytime impairment is expressed in terms of the impact on daytime activities, in cognitive, social, and emotional domains (Léger 2010a). Recent studies have shown that individuals with insomnia exhibit impaired performance in several cognitive functions, including working and episodic memory as well as some aspects of executive functioning (Fortier-Brochu 2012; Bonnet 2014; Shekleton 2014). Cognitive impairments have also been associated with hippocampal atrophy (Joo 2014). Daytime impairments result in reduced productivity at work (Bolge 2009), reduced quality of life (Léger 2012), an increased rate of accidents at work (Kessler 2012) and increased economic burden (Léger 2010b; Rosekind 2010). Indeed, insomnia has major public health implications, representing a significant burden to society. The cumulative expenditures for insomnia include direct costs - such as outpatient encounters and medical procedures, prescription and treatment costs, transportation to and from treatments - and indirect costs - such as healthcare utilisation, lost workplace productivity and increased risk of accidents (Wickwire 2015). Several diseases are correlated with insomnia, particularly hypertension and cardiovascular disorders (Schwartz 1999; Vgontzas 2009a), type 2 diabetes (Vgontzas 2009b) and psychiatric disorders, such as depression and anxiety disorders (Riemann 2007; Baglioni 2011). Drug and alcohol abuse are also associated with insomnia (Johnson 2001; Shibley 2008).

To diagnose insomnia, the most commonly-used instruments are the Diagnostic and Statistical Manual of Mental Disorders (DSM IV-TR, DSM-5), the International Classification of Diseases (ICD-10) and the International Classification of Sleep Disorders (ICDS-2, ICSD-3). Together with disrupted sleep and daytime impairment, the DSM-IV-TR sets a minimum duration of one month to diagnose the condition, and the ICD-10 requires symptom occurrence of at least 3 times a week for a month. In addition, the ICD-10 requires the presence of marked personal distress or interference with personal functioning. Both the ICDS-3 and the DSM-5 establish a duration of at least three months, with a disturbance frequency of at least 3 times per week despite adequate conditions for sleep, and the absence of co-existing other sleep disorders, mental disorders or medical conditions (WHO 1992; APA 2000; AASM 2005; APA 2013; AASM 2014).

Studies which have compared the main diagnostic systems have shown that the diagnosis of insomnia obtained through their use fails to adequately represent the actual extent of the disease in the population (Ohayon 2009). In particular, it has been observed that, in the same population, the diagnoses obtained by the ICD-10 are very low in number compared to those obtained by the DSM-IV-TR or the ICDS-2. Roth 2011 suggested that the DSM-IV-TR was superior to other diagnostic systems. A recent study showed that the prevalence of insomnia diagnoses is estimated to have been reduced by half with the transition from DSM-IV-TR to DSM-5 (Chung 2015). Some researchers identified the measure 'global sleep dissatisfaction' - which considers duration, quality of sleep, or both - as an important element which should be included among diagnostic criteria to improve the accuracy and reliability of the diagnosis (Ohayon 2009; Ohayon 2012).

Description of the intervention

Insomnia treatment is based on sleep hygiene, cognitive behavioural therapy for insomnia (CBT-I) and pharmacological therapy (NICE 2014).

Sleep hygiene refers to a list of behavioural rules designed to increase the likelihood of sleeping well. Poor sleep hygiene can contribute to insomnia, but not cause it; for this reason sleep hygiene education is a necessary, but not a sufficient, treatment for insomnia (Stepanski 2003).

CBT-I is a weekly psychological intervention, normally lasting 8 to 10 weeks. It consists of sleep hygiene instructions, stimulus control therapy, sleep restriction therapy, relaxation and cognitive therapy (Perlis 2005). CBT-I has been shown to be effective in acute treatment as well as in long-term follow-up (Riemann 2009). The combination of CBT-I and pharmacotherapy has been proved to be more effective than CBT-I alone (Morin 2012). CBT-I is often not easily accessible and the prescription of sleep medications is therefore increasing (Ford 2014).

Pharmacotherapy for insomnia consists of different types of drugs. The most commonly-used drugs at present are benzodiazepine receptor agonists (BZRAs) (Ilyas 2012; Ford 2014), which are subdivided into benzodiazepines and benzodiazepine-like drugs (also known as 'Z-drugs'). Other drugs for treating insomnia include antidepressants (mostly tricyclic antidepressants (TCAs)), melatoninergic drugs and orexin receptor antagonists.

BZRAs are positive allosteric modulators at the GABA-A receptor. Gamma-aminobutyric acid (GABA) is the principal inhibitory neurotransmitter of the central nervous system and it is the physiological ligand for GABA-A receptors, which are ligand-gated ion channels. BZRAs' action on the GABA-A receptor is self-limiting, depending on the presence of GABA. In fact, in the absence of GABA, BZRAs cannot open the chloride channel (Rudolph 2011). This self-limiting action is a main reason for the higher safety profile of BZRAs in comparison to previously-used drugs for insomnia, such as barbiturates (Fischbach 1983). Benzodiazepines bind non-selectively onto GABA-A receptors $\alpha 1$, $\alpha 2$, $\alpha 3$, or $\alpha 5$ subunits. Benzodiazepine-like drugs generally have a short half-life, which grants few daytime adverse effects (Rudolph 2011).

TCAs act as inhibitors of serotonin and norepinephrine reuptake and they have also anticholinergic and antihistaminergic properties; it is supposed that the antihistaminergic action is the main reason for their sedating properties (Richelson 1979; Ware 1983).
Melatoninergic drugs are divided into exogenous melatonin and melatonin receptor agonists. Melatonin is an endogenous hormone secreted by the pineal gland and involved in circadian rhythms and the sleep/wake cycle. Exogenous melatonin and melatonin receptor agonists bind MT1 and MT2 melatonin receptors, which regulate the sleep-wake cycle and inhibit arousal signals (Dubocovich 2005; Kato 2005).
Orexin receptor antagonists are a new class of drugs under development. The orexins are excitatory neuropeptides, which convey their actions on two G protein-coupled receptors called OX1 and OX2. They are involved in the sleep/wake cycle, and it has been demonstrated that individuals who suffer from narcolepsy show a deficit of neurons producing orexins (Sakurai 2005).

How the intervention might work

BZRAs are a common choice for insomnia treatment and their use has been systematically increasing over the years (Ford 2014). Benzodiazepines are efficacious in terms of sleep onset latency (SOL) and total sleep time (TST), and they are used clinically for different types of insomnia: short-acting medications are indicated for patients with SOL, while longer-acting medications are preferable for patients with sleep maintenance insomnia (Holbrook 2000; Buscemi 2007). Even though BZRAs have been shown to be effective in the acute treatment of insomnia, they cause important side effects such as cognitive and motor impairments, and somnolence (Holbrook 2000; Poyares 2004; Buscemi 2007; Wafford 2008). In particular, long duration therapies with benzodiazepines may result in the appearance of dependency, withdrawal symptoms (e.g. rebound insomnia) and worsening of sleep parameters (Allison 2003; Poyares 2004). Benzodiazepine-like drugs (i.e. eszopiclone; zaleplon; zolpidem; zopiclone) have a short half-life, produce fewer next-day cognitive and motor impairments (Wafford 2008; Nutt 2010), and at present their clinical use is preferred to benzodiazepines (Wafford 2008; Ford 2014).
Antidepressants are widely used for the treatment of insomnia, and their prescription appears to increase over time together with other non-benzodiazepine drugs (Ford 2014). However, among antidepressants, only doxepin has been approved for the treatment of insomnia and the prescription of other antidepressants (e.g. trazodone, mirtazapine, amytriptiline) is off-label. Doxepin inhibits serotonin and norepinephrine reuptake and inactivates cholinergic, histaminergic and alpha1-adrenergic receptors. At low dose (less than 10mg/day), doxepin has little effect on the serotonergic and adrenergic receptors, promoting sleep onset and duration, and acting as a selective histamine H1 receptor antagonist (Yeung 2015). Therapeutic effects of doxepin are observed at very low dosages (3mg to 6mg/day), improving sleep maintenance without rebound insomnia or physical dependence (Hajak 2001). Common side effects include sedation, nasopharyngitis, gastrointestinal effects, and hypertension (Weber 2010). Doxepin has also been demonstrated to be effective for sleep maintenance and early morning awakenings, which are the most common insomnia-related complaints in the elderly (Krystal 2010).
Melatonin receptor agonists, such as melatonin and ramelteon, have been demonstrated to be a well-tolerated option for the treatment of patients with insomnia characterised by difficulty in sleep onset (Simpson 2008). Ramelteon was associated with reduced subjective sleep latency and improved sleep quality, but not with increased subjective total sleep time. Ramelteon was also associated with improvement in sleep efficiency, and total sleep time by polysomnography, without significant side effects other than somnolence (Kuriyama 2014). At present, no study has demonstrated clear effectiveness for melatoninergic drugs in insomnia.
There are many orexin receptor antagonists under investigation for the treatment of insomnia, and they can be divided into single orexin receptor antagonists (SORAs) and dual orexin receptor antagonists (DORAs) (Equihua 2013). Thus far, the Food and Drug Administration (FDA) has approved only suvorexant, which belongs to the DORAs category, for the treatment of insomnia. In a recent double-blind, placebo-controlled trial, patients undergoing suvorexant therapy showed improved subjective TST and subjective SOL compared with placebo. Those improvements were noticeable after one week of treatment and were maintained throughout one year. The drug was well tolerated by insomnia patients and the most commonly-reported adverse effects were daytime somnolence and fatigue (Michelson 2014).
Other drugs approved for the treatment of insomnia are barbiturates - chloral hydrate, ethchlorvynol, triclofos sodium - but they are not used clinically any longer due to their important adverse effects, toxic effects, and risk of misuse and dependence (Morin 2012; Mowry 2013).
Off-label drugs include antidepressants (with the exception of doxepin which has an indication for the treatment of insomnia) and antipsychotics, which are used for the treatment of insomnia due to psychiatric comorbidities and are considered as a second line treatment (Morin 2012; Saddichha 2010). Antihistamines are still found in many over-the-counter (OTC) sleep aids (Risberg 1975). Most OTCs are non-selective, having anti-muscarinic, anti-adrenergic properties and acting on dopamine and serotonin receptors, which gives rise to unacceptable side effects. Indeed, antihistamines used as sleep-inducing agents can cause drowsiness, but the evidence for their efficacy is very limited and the data on safety and tolerance now discourage their use in insomnia (Morin 2005; Morin 2012; NICE 2014).

Why it is important to do this review

Previous pairwise meta-analyses could not generate clear hierarchies for the efficacy and acceptability of available treatments.

Some Cochrane intervention reviews of sleep drugs for insomnia are in progress (Everitt 2013; Rösner 2013a; Rösner 2013b; Rösner 2013c; Rösner 2013d; Moriichi 2014; Takeshima 2014), while some non-Cochrane systematic reviews have been undertaken (Holbrook 2000; Dündar 2004; Buscemi 2007; Winkler 2014), but this will be the first network meta-analysis in the field. Our intention is to reduce the uncertainty about the efficacy of treatments due to the limited number of direct comparisons as reported in previous standard meta-analyses, and provide an evidence-based hierarchy of the comparative efficacy and acceptability of the drugs approved for the treatment of primary insomnia. This network meta-analysis will help clinicians, patients and policy makers to make informed decisions on the best pharmacological treatments for insomnia. In conclusion, the present review will synthesise the best available clinical evidence, including both direct and indirect comparisons, in order to help clinicians and patients to make informed decisions on the best pharmacological treatments approved for insomnia in terms of efficacy and acceptability.

OBJECTIVES

1) To compare individual pharmacological treatments for insomnia in adults in terms of:

- efficacy, measured as self-rated quality of sleep or satisfaction with sleep; and
- acceptability of treatment.

2) To generate a clinically-useful hierarchy of available pharmacological treatments for insomnia in adults, according to their efficacy and acceptability.

METHODS

Criteria for considering studies for this review

Types of studies

We will include randomised controlled trials (RCTs) comparing active drugs with other active drugs and/or placebo as oral therapy in the treatment of primary insomnia. We will exclude controlled clinical trials, cluster-randomised trials and cross-over trials, in order to avoid possible sources of heterogeneity.

Types of participants

Participant characteristics

Adults aged 18 or older will be included. There will be no limits in terms of gender or ethnicity.

Diagnosis

Insomnia diagnosed according to any standardised diagnostic criteria, such as the DSM-III (APA 1980), DSM-III-R (APA 1987), DSM-IV (APA 1994), DSM-IV TR (APA 2000), DSM-5 (APA 2013), ICD-10 (WHO 1992), International Classification of Sleep Disorders (ICSD) (AASD 1990), ICSD-2 (AASM 2005) or ICSD-3 (AASM 2014).

Co-morbidities

We will include studies on primary insomnia and exclude those considering patients with insomnia due to psychiatric or physical comorbidity. The distinction between primary and secondary insomnia is important for a network meta-analysis, because the severity and the pathophysiologic heterogeneity of the disturbances that cause insomnia are likely to be strong confounders interfering with the reliability of the results. Moreover, the diagnosis has important implications for treatment: therapy for primary insomnia focuses on the improvement of sleep, while therapy for secondary insomnia focuses on the causative medical problem, which also implies that the dose and types of drugs may not be comparable.

Setting

We will consider studies performed in any setting.

Types of interventions

We will include RCTs that evaluate one or more of the following pharmacological interventions as monotherapy, compared to placebo and/or to another active agent:

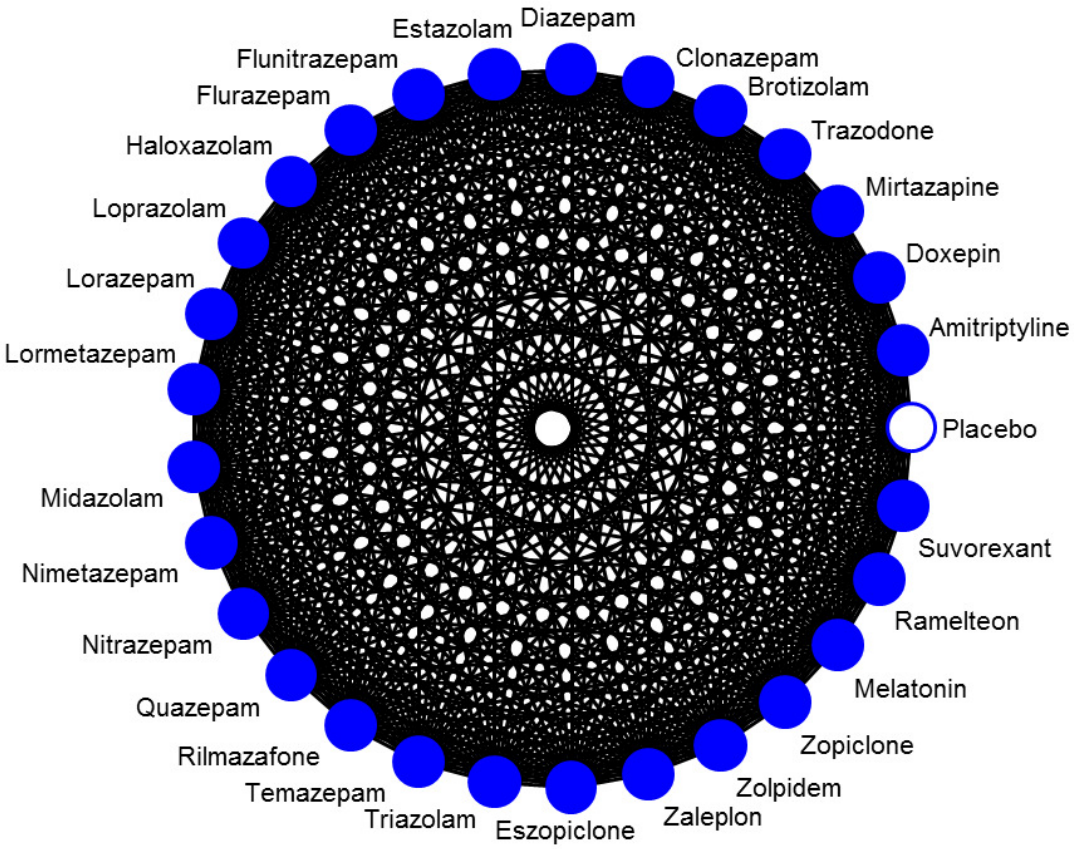
- Antidepressants: amitriptyline; doxepin; mirtazapine; trazodone;
- Benzodiazepines: brotizolam; clonazepam; diazepam; estazolam; flunitrazepam; flurazepam; haloxazolam; loprazolam; lorazepam; lormetazepam; midazolam; nimetazepam; nitrazepam; quazepam; rilmazafone; temazepam; triazolam;
- Benzodiazepine-like agents: eszopiclone; zaleplon; zolpidem; zopiclone;
- Melatoninergic drugs: melatonin; ramelteon;
- Orexin receptor antagonists: suvorexant.

We will exclude barbiturates, chloral hydrate, ethchlorvynol, triclofos sodium and quetiapine due to their important adverse effects, toxic effects, and risk of misuse and dependence (Morin

2012; Mowry 2013). We will also exclude herbal products and medical devices.

Figure 1 shows the network of all possible pairwise comparisons between the eligible treatments. We assume that any patient who meets the inclusion criteria is, in principle, equally likely to be randomised to any of the eligible treatments.

Figure 1. Figure 1 represents all possible pairwise comparisons from the network of the interventions.



Comparability of dosages

We will include only studies randomising patients to drugs within the therapeutic dose. Both fixed-dose and flexible-dose designs will be allowed. We will establish therapeutic doses according to the British National Formulary (BNF) (www.medicinescomplete.com). There is the possibility that some trials may compare one agent at the upper limit of its therapeutic range with another agent at the lower limit of its therapeutic range within the same study. We may look at heterogeneity and then add a binary variable (yes/no) to report if dosages are comparable and use this information for analysis.

We will exclude: (i) combination treatments; (ii) augmentation studies (e.g. drug A+ drug B versus drug A); (iii) all non-pharmacological treatments.

Types of outcome measures

Studies that meet the above inclusion criteria will be included regardless of whether they report on the following outcomes.

Primary outcomes

1. Quality of sleep or satisfaction with sleep index (as a continuous outcome), as measured by any self-rated validated scale, such as the Pittsburgh Sleep Quality Index (PSQI) (Buysse 1989), the Insomnia Severity Index (ISI) (Bastien 2001), or the Leeds Sleep Evaluation Questionnaire (LSEQ) (Parrott 1978; Zisapel 2003). In case other standardised scales were used by some trials, we will use them in the absence of PSQI, ISI or LSEQ.
2. Drop-outs for any reason.

Secondary outcomes

1. Drop-outs due to any adverse event.
2. Daytime functioning as measured by attentional tasks, tests or any self-rated measure of function, for example the 36-item short-form (SF-36) (Ware 1992), the Stanford Sleepiness Scale (SSS) (Hoddes 1973) or the Epworth Sleepiness Scale (ESS) (Johns 1991).
3. Sleep onset latency evaluated by polysomnography, defined as the length of time (in minutes) after lights-out until sleep onset.
4. Wake time after sleep onset (WASO) evaluated by polysomnography, defined as the length of time (in minutes) of wakefulness after the onset of persistent sleep.
5. Total sleep time (TST) evaluated by polysomnography, defined as the total time (in minutes) a person spent sleeping during the in-bed interval. TST is time in bed minus SOL and minus WASO.

Timing of outcome assessment

We will consider outcomes assessed at four weeks post-treatment or at its closest time point. We will include trials with an assessment from one week up to three months. Separately, we will also consider long-term outcomes (more than 3 months).

Hierarchy of outcome measures

For the primary outcome "Quality of sleep" we will select first the PSQI scale firstly, second the ISI scale and third the LSEQ scale. Measures of daytime functioning will be considered and analysed separately.

Search methods for identification of studies

Electronic searches

1. Bibliographic databases

We will search the following bibliographic databases for reports of RCTs using relevant subject headings (controlled vocabularies) and search syntax, appropriate to each resource (Appendix 1):

- Cochrane Central Register of Controlled Trials (CENTRAL, all years)
- Ovid MEDLINE (1950 onwards)
- PubMed (current year)
- Ovid Embase (1980 onwards)
- Ovid PsycINFO (all years)

We will not restrict our search by language, date or publication status.

We will conduct a separate search to identify other systematic reviews and meta-analyses (on Ovid MEDLINE; the Cochrane Database of Systematic Reviews (CDSR); the Database of Abstracts of Reviews of Effects (DARE); and Epistemonikos)

2. International trial registries

We will search the World Health Organization's trials portal (ICTRP) and ClinicalTrials.gov to identify unpublished or ongoing studies.

Searching other resources

Reference lists

We will screen the reference lists of all included studies and relevant systematic reviews to identify additional studies missed from the original electronic searches (for example unpublished or in-press citations).

Correspondence

We will contact trialists and subject experts for information on unpublished or ongoing studies or to request additional trial data.

Data collection and analysis

Selection of studies

Two review authors (FDC, MC) will independently screen titles and abstracts retrieved by the search strategy. Full-texts of potentially relevant studies will then be assessed independently by two

authors (FDC, MC). Disagreements will be resolved through discussion with a third member of the review team (LA).

Data extraction and management

We will use a data collection form to extract study characteristics and outcome data, which has been piloted on at least one study in the review. Two review authors (FDC, MC) will independently extract study characteristics and outcome data from included studies, as follows:

Methods: first author or acronym, year of publication, publication (full-text publication, abstract publication, unpublished data), study design.

Participants: diagnosis, sample size (N), mean age, gender distribution, severity of illness, treatment setting.

Interventions: number of patients allocated to each arm, drug name, dose, route or administration, duration of the interventions and follow-up.

Outcomes: primary and secondary outcomes evaluated.

Adverse events (AEs): AEs as unfavourable symptoms occurring during the course of the study.

Notes: country, funding source; investigational drug versus comparator.

We will note in the 'Characteristics of included studies' table if outcome data were not reported in a usable way. We will resolve disagreements by consensus or by involving a third person (LA). Two review authors (FDC, MC) will enter data into Review Manager (RevMan 2014). We will double-check that data are entered correctly by comparing the data presented in the systematic review with the study reports.

Data on potential effect modifiers

We will extract from each included study data that may act as effect modifiers: age, funding source, studies reported as high risk of bias.

Outcome data

We will extract from each included study:

- self-rated quality of sleep scale, as a continuous outcome: mean and standard deviation (SD);
- drop-outs for any reason: number of participants who dropped out for any reason, of the total number of participants randomised to each arm;
- drop-outs due to any adverse events: number of participants who dropped out because of any adverse event, of the total number of participants randomised to each arm;
- daytime functioning; each single scale will be analysed separately as a continuous outcome: mean and SD;
- polysomnographic outcomes SOL, WASO and TST: mean in minutes and SD.

Assessment of risk of bias in included studies

Two review authors (FDC, MC) will independently assess the risk of bias of each study, using the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We will resolve any disagreements by discussion or by involving another author (LA). The following domains will be assessed: random sequence generation, allocation concealment, blinding of providers and participants, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting. We will judge each potential source of bias as high, low or unclear and provide a supporting quotation from the study report together with a justification for our judgment in the 'Risk of bias' table. We will report the 'Risk of bias' judgements across different studies for each of the domains listed. Where information on risk of bias relates to unpublished data or correspondence with a trialist, we will note this in the 'Risk of bias' table. A judgement of high risk of bias in one or more domain will be considered as a 'high risk' study, a judgement of low risk of bias in all domains will be considered as a 'low risk' study, and a judgement of unclear risk of bias in one or more domains as an 'unclear risk' study. When considering treatment effects, we will take into account the risk of bias for the studies that contribute to that outcome.

Measures of treatment effect

Relative treatment effects

Dichotomous data

Dichotomous outcomes will be analysed by calculating the relative risk (RR) for each trial with the uncertainty in each result being expressed by its 95% confidence interval (CI).

Continuous data

Continuous outcomes will be analysed by calculating the mean difference (MD) with the relative 95% CI when the study used the same instruments for assessing the outcome. We will use the standardised mean difference (SMD) when studies used different instruments.

Relative treatment ranking

For any primary outcome we will also estimate the ranking probabilities for all treatments of being at each possible rank for each intervention. Then we will obtain a treatment hierarchy using the surface under the cumulative ranking curve (SUCRA) and mean ranks. SUCRA will be expressed as a percentage and is interpreted as the percentage of efficacy or safety a treatment achieves in relation to a treatment that would be ranked first without uncertainty (Salanti 2011).

Unit of analysis issues

For simple pairwise meta-analysis, if all arms in a multi-arm trial are to be included in the meta-analysis and one treatment arm is to be included in more than one treatment comparison, then we will divide the number of events and the number of participants in that arm by the number of treatment comparisons made. This method will avoid the multiple use of participants in the pooled estimate of treatment effect, while retaining information from each arm of the trial. It will, however, compromise the precision of the pooled estimate slightly. In the network meta-analysis, we account for the correlation between the effect sizes from multi-arm studies (Higgins 2011, chapter 16.5.4).

Dealing with missing data

We will contact study authors when there are missing or unclear data. If dichotomous outcome data are still missing, they will be managed according to the intention-to-treat (ITT) principle, and we will assume that patients who dropped out after randomisation had a negative outcome. Missing continuous outcome data will either be analysed using the last observation carried forward to the final assessment (LOCF) or, if LOCF data are reported by the trial authors, will be analysed on an endpoint basis, including only participants with a final assessment. When P values, t-values, CIs or standard errors are reported in articles, we will calculate SDs from their values (Furukawa 2006).

Assessment of heterogeneity

In the context of the network meta-analysis, we will assume a common within-network heterogeneity and the generalised Q-statistic estimator will be used for the heterogeneity variance.

Assessment of clinical and methodological heterogeneity within treatment comparisons

To evaluate the presence of heterogeneity deriving from different trial designs or different clinical characteristics of study participants, we will generate descriptive statistics for trial and study population characteristics across all eligible trials that compare each pair of interventions. We will assess the presence of clinical heterogeneity within each pairwise comparison by comparing these characteristics.

Assessment of transitivity across treatment comparisons

We expect that the transitivity assumption will hold assuming that all pairwise comparisons do not differ on average with respect to the distribution of effect modifiers (e.g. age). The assumption of transitivity will be evaluated in each primary outcome by comparing the clinical and methodological characteristics (potential effect modifiers presented in Data extraction and management) across the different pairwise comparisons.

Assessment of reporting biases

The possibility of reporting bias will be evaluated for each primary outcome by means of the contour-enhanced funnel plots if enough studies (at least 10) are available (Peters 2008). These are funnel plots showing areas of statistical significance and they can help to distinguish publication bias from other possible reasons for asymmetry. In a network of interventions each study estimates the relative effect of different interventions, so asymmetry in the funnel plot cannot be judged. To account for this, we will use an adaptation of the funnel plot by subtracting from each study-specific effect size the mean of meta-analysis of the study-specific comparison and plot it against the study's standard error (Chaimani 2012; Chaimani 2013). We will draw the comparison-adjusted funnel plot for all placebo-controlled trials (if at least 10 trials are available). Any asymmetry in the plot indicates the presence of small study effects and not necessarily reporting bias.

Data synthesis

Methods for direct treatment comparisons

We will perform conventional pairwise meta-analyses for primary and secondary outcomes using a random-effects model in RevMan for every treatment comparison with at least two studies (DerSimonian 1986).

Methods for indirect and mixed comparisons

We will perform network meta-analysis (NMA) for primary outcomes. NMA is a method of synthesising information from a network of trials addressing the same question but involving different interventions (Cipriani 2013). NMA combines direct evidence (from studies comparing two treatments, e.g. A versus B) and indirect evidence (e.g. the comparison A versus B comes from studies comparing A and B versus a common comparator C) across a network of randomised trials into a single effect size, and under certain assumptions it can increase the precision in the estimates while randomisation is respected. We will perform NMA using a random-effects model within a frequentist setting assuming equal heterogeneity across all comparisons, and we will account for correlations induced by multi-arm studies (Lu 2006; Salanti 2009). The models will enable us to estimate the probability of each intervention being the best for each outcome, given the relative effect sizes as estimated in NMA. We will perform NMA in Stata 13 using the 'mvmeta' command and self-programmed Stata routines available at http://www.mtm.uoi.gr (Chaimani 2014; White 2011; White 2012).

Results of meta-analysis and NMA will be applied when reasonable and presented as summary relative effect sizes (MD, SMD or RR) for each possible pair of treatments.

Subgroup analysis and investigation of heterogeneity

Measures and tests for heterogeneity

We will statistically assess the presence of heterogeneity for all direct pairwise comparisons using the τ^2. The assessment of statistical heterogeneity in the entire network will be based on the magnitude of the heterogeneity variance parameter (τ^2) estimated from the NMA models. We will compare the magnitude of the heterogeneity variance with the empirical distribution as derived by Turner (Turner 2012). We will also estimate a total I^2 value for heterogeneity in the network as described elsewhere (Jackson 2014).

Assessment of statistical inconsistency

Consistency in a network of treatments refers to the agreement between direct and indirect evidence on the same comparisons. Joint analysis can be misleading if the network is substantially inconsistent. Inconsistency can be present if the trials in the network have very different protocols and their inclusion/exclusion criteria are not comparable or may result as an uneven distribution of the effect modifiers across groups of trials that compare different treatments.

Local approaches for evaluating inconsistency

We will first check for any erroneous data abstraction. Then, to evaluate the presence of inconsistency locally, we will use the loop-specific approach. This method evaluates the consistency assumption in each closed loop of the network separately as the difference between direct and indirect estimates for a specific comparison in the loop (inconsistency factor) (Veroniki 2013). The magnitude of the inconsistency factors and their 95% CIs can then be used to infer as to the presence of inconsistency in each loop. We will assume a common heterogeneity estimate within each loop. We will present the results of this approach graphically in a forest plot using the 'ifplot' command in Stata (Chaimani 2013).

Investigation of heterogeneity and inconsistency

If sufficient studies are available, we will perform network subgroup analyses for the primary efficacy outcome by using age (over 65 years old versus 18 to 65 years old) as possible sources of inconsistency or heterogeneity (Floyd 2000; Ohayon 2002; Ancoli-Israel 2008; Fetveit 2009).

Sensitivity analysis

If enough studies per comparison are identified, we will carry out a sensitivity analysis of the primary outcomes including only trials at low risk of bias in all domains. Moreover, we will perform a sensitivity analysis to assess the robustness of the results if imputations have been applied.

'Summary of findings' table

The main results of the review will be presented in 'Summary of findings' (SoF) tables, as recommended by Cochrane (Schünemann 2011). We will produce the SoF tables for estimates from the NMA based on the methodology developed from the GRADE Working Group (GRADE 2004). For more details, see Salanti 2014. We will include an overall grading of the evidence for the following main outcomes:

Efficacy:

- Self-rated quality of sleep or satisfaction with sleep.

Acceptability:

- Drop-outs for any reason.

We will grade quality of the evidence considering study limitations, indirectness, inconsistency, imprecision of effect estimates, and risk of publication bias. According to the software GRADEpro GDT 2014, we will assign four levels of quality of evidence: high, moderate, low, very low.

ACKNOWLEDGEMENTS

We thank Dr G. Della Marca (Catholic University of the Sacred Heart) for his expert advice.

CRG Funding Acknowledgement:

The National Institute for Health Research (NIHR) is the largest single funder of the Cochrane Common Mental Disorders Group.

Disclaimer:

The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the NIHR, National Health Service (NHS) or the Department of Health.

REFERENCES

Additional references

AASD 1990
American Association of Sleep Disorders. *International Classification of Sleep Disorders: Diagnostic and Coding Manual*. Rochester (MN): American Association of Sleep Disorders, 1990.

AASM 2005
American Academy of Sleep Medicine. *International Classification of Sleep Disorders. Diagnostic and Coding Manual*. 2nd Edition. Westchester: American Academy of Sleep Medicine, 2005.

AASM 2014
American Academy of Sleep Medicine. *International Classification of Sleep Disorders. Diagnostic and Coding Manual*. 3rd Edition. Chicago: American Academy of Sleep Medicine, 2014.

Allison 2003
Allison C, Pratt JA. Neuroadaptive processes in GABAergic and glutamatergic systems in benzodiazepine dependence. *Pharmacology & Therapeutics* 2003;**98**(2):171–95.

Ancoli-Israel 2008
Ancoli-Israel S, Ayalon L, Salzman C. Sleep in the elderly: normal variations and common sleep disorders. *Harvard Review of Psychiatry* 2008;**16**(5):279–86.

APA 1980
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)*. 3rd Edition. Washington, DC: American Psychiatric Association, 1980.

APA 1987
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*. 3rd Edition. Washington, DC: American Psychiatric Association, 1987.

APA 1994
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. 4th Edition. Washington, DC: American Psychiatric Association, 1994.

APA 2000
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. 4th Edition. Washington, DC: American Psychiatric Association, 2000.

APA 2013
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)*. 5th Edition. Washington, DC: American Psychiatric Association, 2013.

Baglioni 2011
Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders* 2011; **135**(1-3):10–19.

Bastien 2001
Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine* 2001;**2**(4):297–307.

Bolge 2009
Bolge SC, Doan JF, Kannan H, Baran RW. Association of insomnia with quality of life, work productivity, and activity impairment. *Quality of Life Research* 2009;**18**(4):415–22.

Bonnet 2014
Bonnet MH, Burton GG, Arand DL. Physiological and medical findings in insomnia: implications for diagnosis and care. *Sleep Medicine Reviews* 2014;**18**(2):111–22.

Buscemi 2007
Buscemi N, Vandermeer B, Friesen C, Bialy L, Tubman M, Ospina M, et al. The efficacy and safety of drug treatments for chronic insomnia in adults: a meta-analysis of RCTs. *Journal of General Internal Medicine* 2007;**22**(9):1335–50.

Buysse 1989
Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research* 1989;**28**(2):193–213.

Chaimani 2012
Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods* 2012;**3**(2): 161–76.

Chaimani 2013
Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in Stata. *PloS One* 2013;**8**(10):e76654.

Chaimani 2014
Chaimani A, Mavridis D, Salanti G. A hands-on practical tutorial on performing meta-analysis with Stata. *Evidence-based Mental Health* 2014;**17**:111–6.

Chung 2015
Chung KF, Yeung WF, Ho FY, Yung KP, Yu YM, Kwok CW. Cross-cultural and comparative epidemiology of insomnia: the Diagnostic and statistical manual (DSM), International classification of diseases (ICD) and International classification of sleep disorders (ICSD). *Sleep Medicine* 2015;**16**(4):477–82.

Cipriani 2013
Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine* 2013;**159**(2):130–7.

DerSimonian 1986
DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;**7**:177–88.

Dubocovich 2005
Dubocovich ML, Markowska M. Functional MT1 and MT2 melatonin receptors in mammals. *Endocrine* 2005;**27** (2):101–10.

Dündar 2004
Dündar Y, Dodd S, Strobl J, Boland A, Dickson R, Walley T. Comparative efficacy of newer hypnotic drugs for the short-term management of insomnia: a systematic review and meta-analysis. *Human Psychopharmacology* 2004;**19**(5): 305–22.

Equihua 2013
Equihua AC, De La Herrán-Arita AK, Drucker-Colin R. Orexin receptor antagonists as therapeutic agents for insomnia. *Frontiers in Pharmacology* 2013;**4**:163.

Everitt 2013
Everitt H, Baldwin DS, Mayers A, Malizia AL, Wilson S. Antidepressants for insomnia. *Cochrane Database of Systematic Reviews* 2013, Issue 10. [DOI: 10.1002/14651858.CD010753]

Fetveit 2009
Fetveit A. Late-life insomnia: a review. *Geriatrics & Gerontology International* 2009;**9**(3):220–34.

Fischbach 1983
Fischbach R. Efficacy and safety of midazolam and vesparax in treatment of sleep disorders. *British Journal of Clinical Pharmacology* 1983;**16**(Suppl 1):S167–71.

Floyd 2000
Floyd JA, Medler SM, Ager JW, Janisse JJ. Age-related changes in initiation and maintenance of sleep: a meta-analysis. *Research in Nursing & Health* 2000;**23**(2):106–17.

Ford 2014
Ford ES, Wheaton AG, Cunningham TJ, Giles WH, Chapman DP, Croft JB. Trends in outpatient visits for insomnia, sleep apnea, and prescriptions for sleep medications among US adults: findings from the National Ambulatory Medical Care survey 1999-2010. *Sleep* 2014; **37**(8):1283–93.

Fortier-Brochu 2012
Fortier-Brochu E, Beaulieu-Bonneau S, Ivers H, Morin CM. Insomnia and daytime cognitive performance: a meta-analysis. *Sleep Medicine Reviews* 2012;**16**(1):83–94.

Furukawa 2006
Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results.. *Journal of Clinical Epidemiology* 2006;**59**:7–10.

GRADE 2004
Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; **328**(7454):1490.

GRADEpro GDT 2014 [Computer program]
GRADE Working Group, McMaster University. GRADEpro GDT. Hamilton (ON): GRADE Working Group, McMaster University, 2014.

Hajak 2001
Hajak G, Rodenbeck A, Voderholzer U, Riemann D, Cohrs S, Hohagen F, et al. Doxepin in the treatment of primary insomnia: a placebo-controlled, double-blind, polysomnographic study. *Journal of Clinical Psychiatry* 2001;**62**(6):453–63.

Higgins 2011
Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane–handbook.org.

Hoddes 1973
Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973;**10**(4):431–6.

Holbrook 2000
Holbrook AM, Crowther R, Lotter A, Cheng C, King D. Meta-analysis of benzodiazepine use in the treatment of insomnia. *Canadian Medical Association Journal* 2000;**162** (2):225–33.

Ilyas 2012
Ilyas S, Moncrieff J. Trends in prescriptions and costs of drugs for mental disorders in England, 1998-2010. *British Journal of Psychiatry* 2012;**200**(5):393–8.

Jackson 2014
Jackson D, Barrett JK, Rice S, White IR, Higgins JP. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Statistics in Medicine* 2014;**33**(21):3639–54.

Johns 1991
Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;**14**(6):540–5.

Johnson 2001
Johnson EO, Breslau N. Sleep problems and substance use in adolescence. *Drug and Alcohol Dependence* 2001;**64**(1): 1–7.

Joo 2014
Joo EY, Kim H, Suh S, Hong SB. Hippocampal substructural vulnerability to sleep disturbance and cognitive impairment in patients with chronic primary insomnia: magnetic resonance imaging morphometry. *Sleep* 2014;**37**(7):1189–98.

Kato 2005
Kato K, Hirai K, Nishiyama K, Uchikawa O, Fukatsu K, Ohkawa S, et al. Neurochemical properties of ramelteon (TAK-375), a selective MT1/MT2 receptor agonist. *Neuropharmacology* 2005;**48**(2):301–10.

Kessler 2012
Kessler RC, Berglund PA, Coulouvrat C, Fitzgerald T, Hajak G, Roth T, et al. Insomnia, comorbidity, and risk of injury among insured Americans: results from the America Insomnia Survey. *Sleep* 2012;**35**(6):825–34.

Krystal 2010
Krystal AD, Durrence HH, Scharf M, Jochelson P, Rogowski R, Ludington E, 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**(11):1553–61.

Kuriyama 2014
Kuriyama A, Honda M, Hayashino Y. Ramelteon for the treatment of insomnia in adults: a systematic review and meta-analysis. *Sleep Medicine* 2014;**15**(4):385–92.

Lu 2006
Lu G, Ades A. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of American Statistical Association* 2006;**101**:447–59.

Léger 2010a
Léger D, Partinen M, Hirshkowitz M, Chokroverty S, Touchette E, Hedner J, EQUINOX (Evaluation of daytime QUality Impairment by Nocturnal awakenings in Outpatients eXperience) survey investigator group. Daytime consequences of insomnia symptoms among outpatients in primary care practice: EQUINOX international survey. *Sleep Medicine* 2010;**11**(10):999–1009.

Léger 2010b
Léger D, Bayon V. Societal costs of insomnia. *Sleep Medicine Reviews* 2010;**14**(6):379–89.

Léger 2012
Léger D, Morin CM, Uchiyama M, Hakimi Z, Cure S, Walsh JK. Chronic insomnia, quality-of-life, and utility scores: comparison with good sleepers in a cross-sectional international survey. *Sleep Medicine* 2012;**13**(1):43–51.

Michelson 2014
Michelson D, Snyder E, Paradis E, Chengan-Liu M, Snavely DB, Hutzelmann J, et al. Safety and efficacy of suvorexant during 1-year treatment of insomnia with subsequent abrupt treatment discontinuation: a phase 3 randomised, double-blind, placebo-controlled trial. *Lancet Neurology* 2014;**13**(5):461–71.

Moriichi 2014
Moriichi A, Tomita N, Sado M, Ota E, Mori R. Interventions for insomnia during pregnancy. *Cochrane Database of Systematic Reviews* 2014, Issue 10. [DOI: 10.1002/14651858.CD011355]

Morin 2005
Morin CM, Koetter U, Bastien C, Ware JC, Wooten V. Valerian-hops combination and diphenhydramine for treating insomnia: a randomized placebo-controlled clinical trial. *Sleep* 2005;**28**(11):1465–71.

Morin 2006
Morin CM, LeBlanc M, Daley M, Gregoire JP, Mérette C. Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. *Sleep Medicine* 2006;**7**(2):123–30.

Morin 2012
Morin CM, Benca R. Chronic insomnia. *Lancet* 2012;**379** (9821):1129–41.

Mowry 2013
Mowry JB, Spyker DA, Cantilena LR Jr, Bailey JE, Ford M. 2012 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 30th Annual Report. *Clinical Toxicology (Philadelphia, Pa.)* 2013;**51**(10):949–1229.

NICE 2014
National Institute for Clinical Excellence. *Insomnia*. London: National Institute for Clinical Excellence, 2014.

Nutt 2010
Nutt DJ, Stahl SM. Searching for perfect sleep: the continuing evolution of GABAA receptor modulators as hypnotics. *Journal of Psychopharmacology (Oxford, England)* 2010;**24**(11):1601–12.

Ohayon 2002
Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Medicine Reviews* 2002;**6**(2):97–111.

Ohayon 2009
Ohayon MM, Reynolds CF 3rd. Epidemiological and clinical relevance of insomnia diagnosis algorithms according to the DSM-IV and the International Classification of Sleep Disorders (ICSD). *Sleep Medicine* 2009;**10**(9):952–60.

Ohayon 2012
Ohayon MM, Riemann D, Morin C, Reynolds CF 3rd. Hierarchy of insomnia criteria based on daytime consequences. *Sleep Medicine* 2012;**13**(1):52–7.

Parrott 1978
Parrott AC, Hindmarch I. Factor analysis of a sleep evaluation questionnaire. *Psychological Medicine* 1978;**8**(2): 325–9.

Perlis 2005
Perlis ML, Jungquist C, Smith MT, Posner D. *Cognitive Behavioral Treatment of Insomnia*. New York: Springer, 2005.

Peters 2008
Peters J, Sutton A, Jones D, Abrams K, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology* 2008;**61**(10):991–6.

Poyares 2004
Poyares D, Guilleminault C, Ohayon MM, Tufik S. Chronic benzodiazepine usage and withdrawal in insomnia patients. *Journal of Psychiatric Research* 2004;**38**(3):327–34.

RevMan 2014 [Computer program]
Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Richelson 1979
Richelson E. Tricyclic antidepressants and histamine H1 receptors. *Mayo Clinic Proceedings* 1979;**54**(10):669–74.

Riemann 2007
Riemann D. Insomnia and comorbid psychiatric disorders. *Sleep Medicine* 2007;**8**(Suppl 4):S15–S20.

Riemann 2009
Riemann D, Perlis ML. The treatments of chronic insomnia: a review of benzodiazepine receptor agonists and psychological and behavioral therapies. *Sleep Medicine Reviews* 2009;**13**(3):205–14.

Risberg 1975
Risberg AM, Risberg J, Ingvar DH. Effects of promethazine on nocturnal sleep in normal man. *Psychopharmacologia* 1975;**43**(3):279–84.

Rosekind 2010
Rosekind MR, Gregory KB. Insomnia risks and costs: health, safety, and quality of life. *American Journal of Managed Care* 2010;**16**(8):617–26.

Roth 2011
Roth T, Coulouvrat C, Hajak G, Lakoma MD, Sampson NA, Shahly V, et al. Prevalence and perceived health associated with insomnia based on DSM-IV-TR; International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; and Research Diagnostic Criteria/International Classification of Sleep Disorders, Second Edition criteria: results from the America Insomnia Survey. *Biological Psychiatry* 2011;**69**(6):592–600.

Rudolph 2011
Rudolph U, Knoflach F. Beyond classical benzodiazepines: novel therapeutic potential of GABAA receptor subtypes. *Nature Reviews. Drug Discovery* 2011;**10**(9):685–97.

Rösner 2013a
Rösner S, Soyka M, Hajak G, Wehrle R, Englbrecht C. Zaleplon for insomnia. *Cochrane Database of Systematic Reviews* 2013, Issue 8. [DOI: 10.1002/14651858.CD010702]

Rösner 2013b
Rösner S, Soyka M, Hajak G, Wehrle R, Englbrecht C. Eszopiclone for insomnia. *Cochrane Database of Systematic Reviews* 2013, Issue 8. [DOI: 10.1002/14651858.CD010703]

Rösner 2013c
Rösner S, Soyka M, Hajak G, Wehrle R, Englbrecht C. Zolpidem for insomnia. *Cochrane Database of Systematic Reviews* 2013, Issue 8. [DOI: 10.1002/14651858.CD010700]

Rösner 2013d
Rösner S, Soyka M, Hajak G, Wehrle R, Englbrecht C. Zopiclone for insomnia. *Cochrane Database of Systematic Reviews* 2013, Issue 8. [DOI: 10.1002/14651858.CD010701]

Saddichha 2010
Saddichha S. Diagnosis and treatment of chronic insomnia. *Annals of Indian Academy of Neurology* 2010;**13**(2):94–102.

Sakurai 2005
Sakurai T. Roles of orexin/hypocretin in regulation of sleep/wakefulness and energy homeostasis. *Sleep Medicine Reviews* 2005;**9**(4):231–41.

Salanti 2009
Salanti G, Marinho V, Higgins J. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009;**62**:857–64.

Salanti 2011
Salanti G, Ades A, Ioannidis J. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 2011;**64**:163–71.

Salanti 2014
Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PloS One* 2014;**9**(7):e99682.

Schwartz 1999
Schwartz S, McDowell Anderson W, Cole SR, Cornoni-Huntley J, Hays JC, Blazer D. Insomnia and heart disease: a review of epidemiologic studies. *Journal of Psychosomatic Research* 1999;**47**(4):313–33.

Schünemann 2011
Schünemann H, Oxman A, Higgins J, Vist G, Glasziou P, Guyatt G. Chapter 11: Presenting results and 'Summary of findings' tables. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions*. 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

Shekleton 2014
Shekleton JA, Flynn-Evans EE, Miller B, Epstein LJ, Kirsch D, Brogna LA, et al. Neurobehavioral performance impairment in insomnia: relationships with self-reported sleep and daytime functioning. *Sleep* 2014;**37**(1):107–16.

Shibley 2008
Shibley HL, Malcolm RJ, Veatch LM. Adolescents with insomnia and substance abuse: consequences and comorbidities. *Journal of Psychiatric Practice* 2008;**14**(3):146–53.

Simpson 2008
Simpson D, Curran MP. Ramelteon: a review of its use in insomnia. *Drugs* 2008;**68**(13):1901–19.

Stepanski 2003
Stepanski EJ, Wyatt JK. Use of sleep hygiene in the treatment of insomnia. *Sleep Medicine Reviews* 2003;**7**(3):215–25.

Takeshima 2014
Takeshima N, Furukawa TA, Hayasaka Y, Ogawa Y, Tajika A. Ramelteon for insomnia. *Cochrane Database of Systematic Reviews* 2014, Issue 3. [DOI: 10.1002/14651858.CD011049]

Turner 2012
Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012;**41**(3):818–27.

Veroniki 2013
Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology* 2013;**42**(1):332–45.

Vgontzas 2009a
Vgontzas AN, Liao D, Bixler EO, Chrousos GP, Vela-Bueno A. Insomnia with objective short sleep duration is associated with a high risk for hypertension. *Sleep* 2009;**32**(4):491–7.

Vgontzas 2009b
Vgontzas AN, Liao D, Pejovic S, Calhoun S, Karataraki M, Bixler EO. Insomnia with objective short sleep duration is associated with type 2 diabetes: a population-based study. *Diabetes Care* 2009;**32**(11):1980–5.

Wafford 2008
Wafford KA, Ebert B. Emerging anti-insomnia drugs: tackling sleeplessness and the quality of wake time. *Nature Reviews. Drug Discovery* 2008;**7**(6):530–40.

Ware 1983
Ware JC. Tricyclic antidepressants in the treatment of insomnia. *Journal of Clinical Psychiatry* 1983;**44**(9 Pt 2): 25–8.

Ware 1992
Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;**30**(6):473–83.

Weber 2010
Weber J, Siddiqui MA, Wagstaff AJ, McCormack PL. Low-dose doxepin: in the treatment of insomnia. *CNS Drugs* 2010;**24**(8):713–20.

White 2011
White IR. Multivariate random-effects meta-regression: updates to mvmeta. *Stata Journal* 2011;**11**:255–70.

White 2012
White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012;**3**(2):111–25.

WHO 1992
World Health Organization. *The Tenth Revision of the International Classification of Diseases and Related Health Problems (ICD-10)*. Geneva: World Health Organization, 1992.

Wickwire 2015
Wickwire EM, Shaya FT, Scharf SM. Health economics of insomnia treatments: the return on investment for a good night's sleep. *Sleep Medicine Reviews* 2015;**28**(30):72–82.

Winkler 2014
Winkler A, Auer C, Doering BK, Rief W. Drug treatment of primary insomnia: a meta-analysis of polysomnographic randomized controlled trials. *CNS Drugs* 2014;**28**(9): 799–816.

Yeung 2015
Yeung WF, Chung KF, Yung KP, Ng TH. Doxepin for insomnia: a systematic review of randomized placebo-controlled trials. *Sleep Medicine Reviews* 2015;**19**:75–83.

Zisapel 2003
Zisapel N, Laudon M. Subjective assessment of the effects of CNS-active drugs on sleep by the Leeds sleep evaluation questionnaire: a review. *Human Psychopharmacology* 2003; **18**(1):1–20. [PUBMED: 12532311]

* *Indicates the major publication for the study*

APPENDICES

Appendix 1. Search strategies

Cochrane Library

Databases on the *Cochrane Library* (CDSR, CENTRAL, DARE, HTA and NHS-EED) will be searched using the following strategy (note specific list of drug terms):
#1 insomni*:ti
#2 MeSH descriptor: [Sleep Initiation and Maintenance Disorders] explode all trees and with qualifier(s): [Drug therapy - DT]
#3 insomni*
#4 sleep near/2 disorder*
#5 MeSH descriptor: [Sleep Disorders] explode all trees and with qualifier(s): [Drug therapy - DT]
#6 MeSH descriptor: [Sleep Initiation and Maintenance Disorders] explode all trees
#7 (#3 or #4 or #5 or #6)
#8 (amitriptyline or doxepin or mirtazapine or trazodone)
#9 (benzodiaz* or brotizolam or clonazepam or diazepam or estazolam or flunitrazepam or flurazepam or haloxazolam or loprazolam or lorazepam or lormetazepam or midazolam or nimetazepam or nitrazepam or oxazepam or quazepam or rilmazafone or temazepam or triazolam)
#10 (nonbenzodiazepin* or non-benzodiazepin* or "Z drug*" or eszopiclon* or zaleplon or zolpidem or zopiclon*)

#11 (melatonin* or ramelteon)
#12 (orexin or suvorexant)
#13 (#8 or #9 or #10 or #11 or #12)
#14 (#7 and #13)
#15 (pharmacotherap* or antidepress* or anti depress* or benzodiazepin* or melaton*):ti
#16 MeSH descriptor: [Antidepressive Agents] explode all trees and with qualifier(s): [Administration & dosage - AD, Therapeutic use - TU]
#17 MeSH descriptor: [Benzodiazepines] explode all trees and with qualifier(s): [Administration & dosage - AD, Therapeutic use - TU]
#18 MeSH descriptor: [Hypnotics and Sedatives] explode all trees and with qualifier(s): [Administration & dosage - AD, Therapeutic use - TU]
#19 (#15 or #16 or #17 or #18)
#20 (#1 or #6) and #19
#21 (#2 or #14 or #20)

OVID databases

Search strategies for MEDLINE, Embase and PsycINFO have been designed to reduce the identification of a large number of irrelevant hits, e.g. treatment studies for depression or other common mental disorders where the symptoms of insomnia are discussed in terms of association, risk or as an adverse effect of psychotropic medication. The searches have also been tailored to exploit the nature and scope of each biomedical database.
We will search OVID MEDLINE (1946 onwards) using the following combination of terms:
1. insomni*.ti,kf.
2. "Sleep Initiation and Maintenance Disorders"/dt [Drug Therapy]
3. Sleep Wake Disorders/dt [Drug Therapy, Therapy]
4. Sleep/dt,th [Drug Therapy]
5. insomni*.ti,ab,kf.
6. (sleep adj2 disorder*).ti,ab,kf.
7. exp "Sleep Initiation and Maintenance Disorders"/
8. (Pittsburgh Sleep Quality Index or PSQI or Insomnia Severity Index or Leeds Sleep Evaluation Questionnaire or LSEQ or (insomnia adj2 (evaluation or index or measure* or questionnaire or rating or scale or schedule or subscale))).ab.
9. or/5-8
10. (amitriptyline or doxepin or mirtazapine or trazodone).mp.
11. (benzodiaz* or brotizolam or clonazepam or diazepam or estazolam or flunitrazepam or flurazepam or haloxazolam or loprazolam or lorazepam or lormetazepam or midazolam or nimetazepam or nitrazepam or oxazepam or quazepam or rilmazafone or temazepam or triazolam).ti,ab,kf,rn,nm.
12. (nonbenzodiazepin* or non-benzodiazepin* or "Z drug*" or eszopiclon* or zaleplon or zolpidem or zopiclon*).mp.
13. (melatonin* or ramelteon).mp.
14. (orexin or suvorexant).mp.
15. or/10-14
16. 9 and 15
17. benzodiazepines/ad,tu or clonazepam/ad,tu or diazepam/ad,tu or flunitrazepam/ad,tu or flurazepam/ad,tu or lorazepam/ad,tu or nitrazepam/ad,tu or temazepam/ad,tu or estazolam/ad,tu or midazolam/ad,tu or triazolam/ad,tu
18. exp Antidepressive Agents/ad,tu,th [Administration & Dosage, Therapeutic Use, Therapy]
19. "Hypnotics and Sedatives"/ad,tu,th [Administration & Dosage, Therapeutic Use, Therapy]
20. (pharmacotherap* or antidepress* or anti depress* or benzodiazepin* or melaton*).ti,kf.
21. or/17-20
22. (1 or 7) and 21
23. controlled clinical trial.pt.
24. randomized controlled trial.pt.
25. (randomi#ed or randomi#ation).ab,ti.
26. randomly.ab.

27. (random* adj3 (administ* or allocat* or assign* or class* or control* or determine* or divide* or distribut* or expose* or fashion or number* or place* or recruit* or subsitut* or treat*)).ab.
28. placebo*.ab,ti.
29. trial.ab,ti.
30. groups.ab.
31. ((singl* or doubl* or tripl* or trebl*) adj3 (blind* or mask* or dummy*)).mp.
32. exp animals/ not (humans.sh. and exp animals/)
33. or/23-31
34. 33 not 32
35. (2 or 3 or 4 or 16 or 22) and 34

OVID Embase will be searched (1980 onwards) using the following combination of terms:
1. *insomnia/
2. insomnia/dt
3. sleep disorder/dt
4. insomni*.ti,kw.
5. exp insomnia/
6. Insomnia Severity Index/
7. insomni*.ti,ab,kw.
8. (Pittsburgh Sleep Quality Index or PSQI or Insomnia Severity Index or Leeds Sleep Evaluation Questionnaire or LSEQ or (insomnia adj2 (evaluation or index or measure* or questionnaire or rating or scale or schedule or subscale))).ab.
9. or/5-8
10. (amitriptyline or doxepin or mirtazapine or trazodone).mp.
11. (benzodiaz* or brotizolam or clonazepam or diazepam or estazolam or flunitrazepam or flurazepam or haloxazolam or loprazolam or lorazepam or lormetazepam or midazolam or nimetazepam or nitrazepam or oxazepam or quazepam or rilmazafone or temazepam or triazolam).ti,ab,kw,rn.
12. brotizolam/ or clonazepam/dt or diazepam/dt or estazolam/ or flunitrazepam/dt or flurazepam/dt or haloxazolam/ or loprazolam/ or lorazepam/dt or lormetazepam/ or midazolam/dt or nimetazepam/ or quazepam/ or rilmazafone/ or temazepam/ or triazolam/dt
13. *clonazepam/ or *diazepam/ or *flunitrazepam/ or *flurazepam/ or *lorazepam/ or *midazolam/ or *temazepam/ or *triazolam/
14. (nonbenzodiazepin* or non-benzodiazepin* or "Z drug*" or eszopiclon* or zaleplon or zolpidem or zopiclon*).mp.
15. eszopiclone/dt or zopiclone/dt or zolpidem/dt or zolpidem tartrate/dt or zaleplon/dt
16. *eszopiclone/ or *zopiclone/ or *zolpidem/ or *zolpidem tartrate/ or *zaleplon/
17. (melatonin* or ramelteon).mp.
18. (orexin or suvorexant).mp.
19. or/10-18
20. randomized controlled trial.de.
21. randomization.de.
22. placebo.de.
23. placebo$.ti,ab.
24. randomi#ed.ti,ab.
25. randomly.ab.
26. ((singl$ or doubl$ or trebl$ or tripl$) adj3 (blind$ or mask$ or dummy)).mp.
27. ((animal or nonhuman) not (human and (animal or nonhuman))).de.
28. or/20-26
29. 28 not 27
30. 9 and 19 and 29
31. limit 30 to exclude medline journals
32. (1 or 2 or 3 or 4) and 19 and 29
33. 33 or 34

OVID PsycINFO will be searched (all years) using the following combination of terms:
1. insomnia/
2. insomni*.ti,ab,id.

3. (Pittsburgh Sleep Quality Index or PSQI or Insomnia Severity Index or ISI or Leeds Sleep Evaluation Questionnaire or LSEQ or (insomnia adj2 (evaluation or index or measure* or questionnaire or rating or scale or schedule or subscale))).tm.
4. or/1-3
5. sleep disorders/
6. (insomni* or (sleep adj2 disorder*)).ti,id.
7. "3340".cc. *[Classification Code=Clinical Psychopharmacology]*
8. ((1 or 5 or 6) and 7)
9. (amitriptyline or doxepin or mirtazapine or trazodone).ti,ab,id,sh.
10. (benzodiaz* or brotizolam or clonazepam or diazepam or estazolam or flunitrazepam or flurazepam or haloxazolam or loprazolam or lorazepam or lormetazepam or midazolam or nimetazepam or nitrazepam or oxazepam or quazepam or rilmazafone or temazepam or triazolam).ti,ab,id,sh.
11. (nonbenzodiazepin* or non-benzodiazepin* or "Z drug*" or eszopiclon* or zaleplon or zolpidem or zopiclon*).ti,ab,id,sh.
12. (melatonin* or ramelteon).ti,ab,id,sh.
13. (orexin or suvorexant).ti,ab,id,sh.
14. or/9-13
15. (pharmacotherap* or antidepress* or anti depress* or benzodiazepin* or melaton*).ti.
16. exp antidepressant drugs/
17. exp benzodiazepines/
18. exp hypnotic drugs/
19. drug therapy/
20. or/15-19
21. treatment effectiveness evaluation.sh.
22. clinical trials.sh.
23. placebo.sh.
24. placebo.ti,ab,id.
25. randomly.ab.
26. randomi#ed.ti,ab,id.
27. (random* adj3 (administ* or allocat* or assign* or class* or control* or determine* or divide* or distribut* or expose* or fashion or number* or place* or recruit* or subsitut* or treat*)).ab.
28. trial.ti,ab,id.
29. ((singl* or doubl* or trebl* or tripl*) adj3 (blind* or mask* or dummy)).ti,ab,id.
30. (control* adj3 group*).ab.
31. "2000".md. *[Methodology=Treatment Outcome/Clinical Trial]*
32. or/21-31
33. (4 and 14 and 32)
34. (8 and 32)
35. (6 and 20 and 32)
36. or/33-35

PubMed will be searched (current year only) using the following terms:
((insomni* OR sleep disorder* OR sleep wake disorder* OR sleep initiation OR sleep onset OR sleep latency) AND (amitriptyline OR doxepin OR mirtazapine OR trazodone OR benzodiazepin* OR brotizolam OR clonazepam OR diazepam OR estazolam OR flunitrazepam OR flurazepam OR haloxazolam OR loprazolam OR lorazepam OR lormetazepam OR midazolam OR nimetazepam OR nitrazepam OR oxazepam OR quazepam OR rilmazafone OR temazepam OR triazolam OR nonbenzodiazepin* OR non-benzodiazepin* OR "Z drug" OR "Z drugs" OR eszopiclon* OR zaleplon OR zolpidem OR zopiclon* OR melatonin* or ramelteon OR orexin OR suvorexant) AND (random* OR placebo OR trial OR (control* AND group) OR treat*[Title] OR efficacy[Title] OR effectiveness[Title)) Filters: Publication date from 2015/01/01 to 2016/01/31

International Trial Registries:

The World Health Organisations trials portal (ICTRP) will be searched using the following search string:
(insomnia and amitriptyline or insomnia and doxepin or insomnia and mirtazapine or insomnia and trazodone or insomnia and brotizolam or insomnia and clonazepam or insomnia and diazepam or insomnia and estazolam or insomnia and flunitrazepam or insomnia and flurazepam or insomnia and haloxazolam or insomnia and loprazolam or insomnia and lorazepam or insomnia and lormetazepam or insomnia and midazolam or insomnia and nimetazepam or insomnia and nitrazepam or insomnia and quazepam or insomnia and rilmazafone or insomnia

and temazepam or insomnia and triazolam or insomnia and eszopiclone or insomnia and zaleplon or insomnia and zolpidem or insomnia and zopiclone or insomnia and melatonin or insomnia and ramelteon or insomnia and orexin or insomnia and suvorexant)
An advanced search of ClinicalTrials.gov will be conducted, for:
CONDITION=*insomnia* and STUDY TYPE=*interventional*
Results will be imported into Excel and filtered by INTERVENTION=*drug* and STUDY DESIGNS=*randomized*

WHAT'S NEW

Date	Event	Description
26 September 2016	Amended	Contact details updated.

CONTRIBUTIONS OF AUTHORS

All of the authors actively contributed to the development of this protocol.

DECLARATIONS OF INTEREST

Franco De Crescenzo, Francesca Foti, Marco Ciabattini, Cinzia Del Giovane, Monica Sañé Schepisi, Digby Quested, Andrea Cipriani, Corrado Barbui, Laura Amato have no conflict of interest to declare relating to this work.

Norio Watanabe has research funds from the Japanese Ministry of Health Labor and Welfare and the Japanese Ministry of Education, Science, and Technology. He has also received royalties from Sogensha and Paquet, and speaking fees and research funds from Asahi Kasei, Dai-Nippon Sumitomo, Eli Lilly, GlaxoSmithKline, Janssen, Meiji, MSD, Otsuka and Pfizer during the last five years.

Sleep Biol. Rhythms (2016) 14:231–239
DOI 10.1007/s41105-016-0055-9

REVIEW ARTICLE

The association between shift work and health: a review

Osamu Itani[1] · **Yoshitaka Kaneita**[1]

Received: 13 December 2015 / Accepted: 8 February 2016 / Published online: 25 February 2016

Abstract We examined previously published systematic review studies on associations between shift work and health, using the PubMed (MEDLINE) database for our literature search. We eventually selected 30 studies: 2 studies on sleep disorders, 7 on cancer, 7 on metabolic endocrine disorders, 7 on reproduction, 4 on cardiovascular disease, 1 on gastrointestinal disorders, and 2 on the types of shift work and health-related outcomes. Meta-analyses based on quantitative combination of the data from these studies showed that shift work significantly increased the risk of the following disorders: breast cancer, diabetes mellitus, preterm delivery, abortion, low birth weight, small-for-gestational-age infants, menstrual disruption, infertility, ischemic heart disease, and ischemic stroke. Some previous studies had also reported significantly increased risks of sleep disturbance, prostate cancer, body weight change, metabolic syndrome, and fertility and gastrointestinal disorders. Some studies had gathered substantial data, including those obtained from meta-analyses, which indicated significant associations. In contrast, other studies were unable to present sufficient evidence because of the smaller number of data sets included. Therefore, based on these findings, further accumulation of epidemiological studies on this theme is warranted.

Keywords Night work · Literature review · Cancer · Metabolic · Reproduction · Cardiovascular

Introduction

In developed countries, various factors such as industrialization, urbanization, and widespread utilization of information technologies have resulted in an increase in round-the-clock operations at industrial plants and for services. In such an environment, the demand for various patterns of shift work has increased. The International Labour Organization (ILO) defines shift work as "a method of organization of working time in which workers succeed one another at the workplace so that the establishment can operate longer than the hours of work of individual workers [1]." The proportion of night workers, including shift workers, has been reported to be 17.7–25.9 % in the United States [2] and 17.3 % in 27 European Union member states [3]. In Japan, the proportion of night workers has increased from 13.3 % in 1997 to 21.8 % in 2012 [4].

As the demand for shift work has increased, its effect on health has become a growing concern. It has been suggested that the derangement of circadian rhythm caused by shift work adversely affects both physical and mental health [5], and many epidemiological studies have investigated this issue. In particular, since the 2000s, using techniques such as systematic reviews and meta-analyses, strong evidence has been presented based on accumulated data from previous studies. In this review article, we present all of the results from previous systematic review studies on shift work and health outcomes, and review the evidence obtained.

✉ Yoshitaka Kaneita
nusmpublichealth@gmail.com

[1] Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, 1-1 Idaigaoka, Hasama-machi, Yufu City, Oita 879-5593, Japan

Materials and methods

Search strategy

The literature review was performed by searching the electronic database PubMed (MEDLINE). The search terms were shift work, night work, systematic review (retrieval style: shift work[Title/Abstract] OR night work[Title/Abstract]) AND systematic review[Title/Abstract] AND English[lang]). As a result of our search, 48 articles were identified. After confirming the content of these 48 articles, we eventually selected 30 studies.

Systematic reviews on shift work and health

Shift work and sleep disorder

A number of systematic reviews have investigated the associations between shift work and sleep disorders, and between shift work and sleep quality (Table 1). Niu et al. [6] performed a systematic review on shift work and sleep quality in 2011, selecting 9 studies. Their evidence levels were reported to range from Level III-2 to Level IV, as measured by the 6-point Quality Scale Assessment.

In 2015, Linton et al. [7] performed a systematic review on the associations between shift work and sleep disturbance and between night work and sleep disturbance. Six studies were selected for shift work and 3 for night work. A meta-analysis of shift work indicated no significant associations, with an odds ratio (OR) of 1.17 (95 % confidence interval [CI] 0.96–1.43).

Shift work and cancer

Among systematic reviews on the association between shift work and various types of cancer (Table 2), Megdal et al. [8] investigated the association between night work and breast cancer risk based on 13 selected studies and a meta-analysis and found a significant increase in breast cancer risk, with a standardized incidence ratio of 1.44 (95 % CI 1.26–1.65).

In 2008, Kolstad performed a systematic review on associations between nightshift work or shift work and the risk of developing breast cancer and other cancers (including prostate cancer, colon cancer, and all cancers). Eight studies focusing on breast cancer, 3 on prostate cancer, 3 on colon cancer, and 4 on all types of cancer were selected [9]. The authors concluded, "There is limited evidence for a causal association between nightshift work and breast cancer, while there is insufficient evidence for prostate cancer, colon cancer, and overall cancer."

In 2012, Sigurdardottir et al. [10] performed a systematic review on the association between circadian rhythm disruption and prostate cancer risk. They selected 4 studies on shift work, of which 3 reported that shift work increased the risk of developing prostate cancer.

In 2013, Ijaz et al. [11] performed a systematic review on the association between nightshift work and breast cancer development. Twelve case–control studies and 4 cohort studies were selected and meta-analyses were performed for both types of study. The results for the case–control studies indicated a significant increase in the risk ratio (RR) (RR 1.09, 95 % CI 1.02–1.20). The RR for the cohort studies was 1.01 (95 % CI 0.97–1.05).

In the same year, Jia et al. [12] also performed a systematic review on the association between night work and breast cancer risk, selecting 8 case–control studies and 5 cohort studies. Meta-analyses were performed for both types of study. The results for the case–control studies indicated a significant increase in the RR (RR 1.32, 95 % CI 1.17–1.50). The RR for the cohort studies was 1.08 (95 % CI 0.97–1.21), and thus no significant increase was recognized.

In 2013, Kamdar et al. [13] also performed a systematic review on the association between nightshift work and the risk of breast cancer. Fifteen studies were selected and meta-analyses were performed based on the duration of the worker's engagement in nightshift work. The results indicated that the RR for ever nightshift workers was 1.21

Table 1 Systematic reviews of shift work and sleep disorder

References	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Niu et al. [6]	Sleep Quality	SCOPUS, OVID, Blackwell Science, EBSCO Host, PsycINFO, Cochrane Controlled Trials Register, CEPS	9	(−)
Linton et al. [7]	Sleep disturbances	PubMed, Embase, PsycInfo, Cochrane library, NIOSHTIC-2	Shift work: 6(2) Night work: 3(0)	Shift work; OR 1.17 (95 % CI 0.96–1.43)

OR odds ratio, *CI* confidential interval

Table 2 Systematic reviews of shift work and cancer

References	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Megdal et al. [8]	Breast cancer	MEDLINE	13(8)	RR 1.48 (95 % CI 1.36–1.61)
Kolstad [9]	Breast cancer	MEDLINE Science Citation Index	8	(−)
	Prostate cancer		3	
	Colon cancer		3	
	All cancers		4	
Sigurdardottir et al. [10]	Prostate cancer	PubMed	4(2)	(−)
Ijaz et al. [11]	Breast cancer	EMBASE, CINAHL, PsycInfo, LILACS, OSH Update, ProQuest	Case control: 12(4)	RR 1.09 (95 % CI 1.02–1.20)
			Cohort: 4(0)	RR 1.01 (95 % CI 0.97–1.05)
Jia et al. [12]	Breast Cancer	PubMed, EMBASE, CNKI, Chinese Wanfang Database	case control: 8	RR 1.32 (95 % CI 1.17–1.50)
			Cohort: 5	RR 1.08 (95 % CI 0.97–1.21)
Kamdar et al. [13]	Breast Cancer	MEDLINE, EMBASE, CINAHL, Web of Science, ProQuest	Ever: 8	RR 1.21 (95 % CI 1.00–1.47)
			Short-term (<8 years): 13	RR 1.13 (95 % CI 0.97–1.32)
			long terms (≥8 years): 9	RR 1.04 (95 % CI 0.92–1.18)
Wang et al. [14]	Breast Cancer	MEDLINE, EMBASE, PSYCInfo, APC Journal Club, Global Health	10(5)	RR 1.19 (95 % CI 1.05–1.35)

RR relative risk, *CI* cofidential interval

(95 % CI 1.00–1.47), that for short-term nightshift workers (<8 years) was 1.13 (95 % CI 0.97–1.32), and that for long-term nightshift workers (≥8 years) was 1.04 (95 % CI 0.92–1.18). Thus, none of the findings were significant. Therefore the authors concluded that there was only weak evidence to support the findings of previous studies indicating that nightshift work was associated with an increased risk of breast cancer.

Wang et al. [14] also performed a systematic review on the association between nightshift work and the risk of breast cancer, selecting 10 studies. A meta-analysis of all these studies indicated a significant increase in the risk of breast cancer development (RR 1.19, 95 % CI 1.05–1.35). This review also analyzed the dose–response relationship between nightshift work and breast cancer. It was concluded that every 5-year increase of nightshift work exposure enhanced the risk of breast cancer in women by 3 % (pooled RR 1.03, 95 % CI 1.01–1.05) and that an increase of 500 night shifts would result in a 13 % (RR 1.13, 95 % CI 1.07–1.21) increase in the risk of breast cancer. It is considered that one of the reasons for the differences in the results of the research performed by Kamdar et al. and Wang et al. is the difference in the statistical analysis method that they each used. In Kamdar et al. [13] the sample was split into two groups, those who had been carrying out shift work for 8 years or more, and those who had been carrying out shift work for less than 8 years, and a meta-analysis of those groups performed. In contrast with this, Wang et al. [14] carried out their analysis regarding the increase in breast cancer risk according to the sum of both the frequency of, and number of years shift work was carried out. Taking into account the fact that in this way differences in statistical analysis method produce differences in the results what is desired from here on is an accumulation of evidence, in addition to the introduction of more advanced statistical analysis techniques.

Shift work and metabolic endocrine disorders

Among systematic reviews on associations between shift work and various types of metabolic endocrine disorders (Table 3). Niu et al. [6] performed a systematic review on the association between shift rotation and the cortisol profile. They analyzed 5 studies of the effects of shift work on cortisol levels, and concluded that cortisol secretion was increased in shift workers who slept during the day.

In 2011, van Drongelen et al. [15] performed a systematic review of longitudinal studies on the effects of shift work on body weight change. Eight studies were selected, of which 7 indicated an association between shift work exposure and a change in body weight. The authors concluded, "There was strong evidence for a crude relationship between shift work and body weight increase." However, they also stated, "The evidence for a confounders-adjusted relationship between shift work exposure and body weight was considered to be insufficient."

Table 3 Systematic reviews of shift work and endcrine disorder

References	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Niu et al. [6]	Cortisol profile, sleep quality, fatigue, attention level	SCOPUS, OVID, Blackwell Science, EBSCO Host, PsycINFO, Cochrane controlled trials register, CEPS	5	(−)
van Drongelen et al. [15]	Body weight change	MEDLINE, EMBASE, the Cochrane library, PsycINFO	8(7)	(−)
Canuto et al. [16]	Metabolic syndrome	PubMed, EMBASE, Web of Science, Science Direct	10(8)	(−)
Staufenbiel et al. [17]	Hair cortisol	Web of Knowledge, PubMed	1(1)	(−)
Knutsson and Kempe [19]	Diabetes	PubMed, Noishtic2, Science Direct	5(3)	(−)
Gan et al. [20]	diabetes mellitus	PubMed, EMBASE, Web of Science ProQuest Dissertation	Cohort: 8 Cross-sectional: 4	Cohort; OR 1.12 (95 % CI 1.06–1.19) Cross-sectional; OR 1.06 (95 % CI 1.03–1.09)
Ulhoa et al. [21]	Cortisol, Obesity, insulin resistance, diabetes, metabolic syndrome	PubMed	Cortisol: 16 Obesity: 15 Insulin resistance, diabetes, metabolic syndrome: 15	(−)

OR odds ratio, *CI* confidential interval

In 2013, Canuto et al. [16] performed a systematic review on the association between shift work and metabolic syndrome, selecting 10 studies. They reported that 8 of those studies had found a positive association between shift work and metabolic syndrome after controlling for socio-demographic and behavioral factors.

In 2013, Staufenbiel et al. [17] performed a systematic review on hair cortisol, stress exposure, and mental health, and selected one study on shift work and hair cortisol. This one study performed by Manenschijn et al. [18] in 2011, and they measured the level of hair cortisol in 33 shift workers and 89 day workers, and reported that the shift workers had a significantly higher level of hair cortisol.

In 2014, Knutsson and Kempe [19] performed a systematic review on the association between shift work and diabetes mellitus. Five studies were selected, of which 4 reported that the risk of diabetes mellitus was significantly increased by shift work.

In 2015, Gan et al. [20] also performed a systematic review on the association between shift work and diabetes mellitus. Twelve studies were selected and a meta-analysis was performed. The results indicated a significant increase in diabetes risk with shift work (OR 1.09, 95 % CI 1.05–1.12). In addition, subgroup analyses suggested a stronger association for men (OR 1.37, 95 % CI 1.20–1.56) than for women (OR 1.09, 95 % CI 1.04–1.14). They also reported that all shift work schedules, with the exception of mixed shifts and evening shifts, had been associated with a significantly higher risk of diabetes mellitus than normal daytime schedules, and that the difference among those shift work schedules had been significant.

In 2015, Ulhoa et al. [21] performed a systematic review on the association between shift work and endocrine disorders (cortisol levels, the presence of obesity, insulin resistance, diabetes, or metabolic syndrome). Fifteen studies on obesity, 8 on diabetes mellitus, 3 on metabolic syndrome, 16 on cortisol levels, and 4 on insulin resistance were selected. They concluded that the epidemiologic evidence showed that shift and night work were associated with an elevated risk of metabolic disorders.

Shift work and reproduction

Among systematic reviews on associations between shift work and reproduction (Table 4). Bonzini et al. [22] reviewed the association of shift work with the risk of premature birth, low birth weight (LBW), and pre-

Table 4 Systematic reviews of shift work and reproduction

References	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Bonzini et al. [22]	Preterm delivery, being small-for-gestational age at delivery, pre-eclampsia and pregnancy-induced hypertension	MEDLINE, EMBASE	Preterm delivery: 14(5) being small-for-gestational age at delivery: 6(1) pre-eclampsia and pregnancy-induced hypertension: 2(0)	Preterm delivery; RR 1.31 (95 % CI 1.16–1.47) being small-for-gestational age at delivery; RR 1.07 (95 % CI 0.96–1.19)
Quansah and Jaakkola [23]	Spontaneous abortion among Nurses	PubMed, EMBASE	4	OR 1.44 (95 % CI 1.17–1.39)
Bonzini et al. [24]	preterm delivery (PTD), low birthweight (LBW), small-for-gestational-age infants (SGA), pre-eclampsia	MEDLINE	23	PTD; RR 1.16 (95 % CI 1.00–1.33) LBW; RR 1.27(95 % CI 0.93–1.74) SGA; RR 1.12 (95 % CI 1.03–1.22)
Bonde et al. [25]	miscarriage	MEDLINE, EMBASE	Shift work: 13 Fixed night work: 6	Shift work; OR 1.12 (95 % CI 0.96–1.30) Fixed night work; OR 1.51 (95 % CI 1.27–1.75)
Chau et al. [26]	pregnancy, fertility, menstrual cycles	CINAHL, MEDLINE, Sociological Abstracts, and Business Source Premier for primary research studies	Pregnancy: 13 Fertility: 3(2) Menstrual cycles: 4(3)	(−)
Stocker et al. [27]	Menstrual disruption, infertility, early spontaneous pregnancy loss	MEDLINE, EMBASE, Google Scholar, the Cochrane Library, online publications of national colleges, the ClinicalTrials.gov	Menstrual disruption: 4 infertility: 4 Early spontaneous pregnancy loss: 7	Menstrual disruption; OR 1.22 (95 % CI 1.15–1.29) Infertility; OR 1.80 (95 % CI 1.01–3.20) Early spontaneous pregnancy loss; OR 0.96 (95 % CI 0.88–1.05)
van Melick et al. [28]	preterm birth	MEDLINE, EMBASE	13(0)	OR 1.04 (95 % CI 0.90–1.20)

RR relative risk, *CI* cofidential interval, *OR* odds ratio

eclampsia. Fourteen studies on preterm delivery (PTD), 6 on small-for-gestational age delivery, and 2 on pre-eclampsia and pregnancy-induced hypertension were selected. The result of the meta-analyses indicated a significant increase in the risk of PTD with shift work (RR 1.31, 95 % CI 1.16–1.47). The RR for being small for gestational age at delivery was 1.07 (95 % CI 0.96–1.19).

In 2010, Quansah and Jaakkola [23] performed a systematic review on the association between shift work and spontaneous abortion. Four studies were selected and a meta-analysis was performed. The results indicated that there was a significant increase in the risk of spontaneous abortion with shift work (OR 1.28, 95 % CI 1.17–1.39).

In 2011, Bonzini et al. [24] performed a systematic review on associations between shift work and pregnancy outcomes [PTD, LBW, small-for-gestational-age infants (SGA), and pre-eclampsia]. Twenty-three studies were selected and meta-analyses were performed. The results indicated that the RR was 1.16 for PTD (95 % CI 1.00–1.33), 1.27 for LBW (95 % CI 0.93–1.74), and 1.12 for SGA (95 % CI 1.03–1.22).

In 2013, Bonde et al. [25] performed a systematic review on the association between shift work and the risk of miscarriage. Thirteen studies on shift work and 6 on fixed night work were selected, and meta-analyses were performed. The results indicated no significant association between shift work and miscarriage (OR 1.12, 95 % CI 0.96–1.30), but a significant increase in the risk of miscarriage was recognized for fixed night work (OR 1.51, 95 % CI 1.27–1.75).

In 2014, Chau et al. [26] performed a systematic review on the association between night work and the reproductive health of women (pregnancy, fertility, and menstrual cycles). Thirteen studies on pregnancy, 3 on fertility, and 4 on menstrual cycles were selected. In the Discussion section of the review, they stated, "Evidence of the impact of night work on female reproductive health as presented in the current literature is inconclusive."

In 2014, Stocker et al. [27] performed a systematic review on the association between shift work and early reproductive outcomes (menstrual disruption, infertility, and early spontaneous pregnancy loss). Four studies on menstrual disruption, 4 on fertility, and 7 on early spontaneous pregnancy loss were selected, and meta-analyses were performed for each of them. The results indicated that there was a significant increase in the risk of menstrual disruption (OR 1.22, 95 % CI 1.15–1.29) and infertility (OR 1.80, 95 % CI 1.01–3.20) with shift work. The OR for early spontaneous pregnancy loss was 0.96 (95 % CI 0.88–1.05).

In 2014, van Melick et al. [28] performed a systematic review on the association between shift work and preterm birth. Thirteen studies were selected and a meta-analysis was performed. The results indicated no significant associations (OR 1.04, 95 % CI 0.90–1.20).

Shift work and cardiovascular disease

Among systematic reviews on associations between shift work and cardiovascular disease (Table 5), Frost et al. [29] performed a systematic review on shift work and the risk of ischemic heart disease. Sixteen studies were selected and a meta-analysis was performed. The results indicated a significant increase in the risk of ischemic heart disease with shift work (RR 1.48, 95 % CI 1.36–1.61).

Togo et al. [30] performed a systematic review on the association between shift work and heart rate variability. They selected 11 studies and reported that several of them had shown altered circadian rhythms of heart rate variability in night shift workers.

In 2012, Hwang et al. [31] performed a systematic review on the association between shift work and cardiovascular disease. Four studies were selected, of which 3 reported a significant increase in the risk of cardiovascular disease with shift work.

Vyas et al. [32] performed a systematic review on the association between shift work and vascular events (myocardial infarction, ischemic stroke, and coronary events). Thirty-four studies were selected and meta-analyses were performed for each event. The results indicated significant increases in the risk of myocardial infarction (RR 1.23, 95 % CI 1.15–1.31), ischemic stroke (RR 1.05, 95 % CI 1.01–1.09), and coronary events (RR 1.24, 95 % CI 1.10–1.39) with shift work.

Shift work and gastrointestinal disorders

Among systematic reviews on associations between shift work and gastrointestinal disorders (Table 6). Knutsson and Boggild [33] performed a systematic review on shift work and gastrointestinal disorders. Six studies on gastrointestinal symptoms, 6 on peptic ulcer disease, and 3 on functional gastrointestinal disease were selected. Of these studies, 4 on gastrointestinal symptoms, 5 on peptic ulcer disease, and 2 on functional gastrointestinal disease reported significant increases in the risk of such disorders with shift work.

Types of shift work and health-related outcomes

Among systematic reviews on associations between types of shift work and health-related outcomes (Table 7). Bambra et al. [34] published a systematic review on the use of intervention strategies, which involved redesigning shift work schedules and their effects on health. Using 27 electronic databases, a literature search of the effects of reorganizing shift work schedules on health and work-life balance was performed, and 26 studies were selected and evaluated based on the types of intervention. Among the examined intervention types (speed of rotation, direction of rotation, removal of shift work rotation, changes to night work, later start and finish times, changes to weekend working, decreased shift lengths, and self-scheduling), switching from slow to fast rotation, changing from backward to forward rotation, and self-scheduling of shifts were reported to have beneficial effects.

In 2015, Vedaa et al. [35] published a systematic review on quick returns in rotating shift work and health-related outcomes. They defined 11.0 h or less between 2 consecutive shifts as a quick return. Next, they performed a systematic literature search on outcome measures of health, sleep functional ability, and work-life balance, and selected 22 studies. They concluded that quick returns had detrimental effects on sleep, sleepiness, and fatigue. However, they also stated that the evidence for the effects on more chronic outcomes (physical and mental health, and work-life balance) was inconclusive.

Biological mechanisms underlying the occurrence of health problems caused by shift work

Faraut et al. [36] analyzed the biological mechanisms underlying the occurrence of health problems caused by shift work. Shift workers suffer from sleep restriction or circadian misalignment, although there may be certain individual differences (based on age, attributes, and sleep characteristics) [37]. Experimental circadian misalignment and sleep restriction may induce neuroendocrine stress [38, 39], immune responses [40–44], or oxidative stress [45–47]. Moreover, neuroendocrine stress may increase the secretion of cortisol and catecholamines, leading to hypertension that may eventually increase the onset risk of cardiovascular diseases [48–51]. Immune responses may increase leukocytes [40, 41], inflammatory cytokines [44], and C-reactive protein [43]. This can lead to inflammation,

Table 5 Systematic reviews of shift work and cardiovascular disease

References	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Frost et al. [29]	ischemic heart disease	MEDLINE	16	RR 1.48 (95 % CI 1.36–1.61)
Togo and Takahashi [30]	Heart rate Variability	MEDLINE (Ovid, PubMed)	11	(−)
Hwang and Hong [31]	Cardiovascular disease	PsycINFO, Educational Resource Information Center, PubMed, CINAHL	4(3)	(−)
Vyas et al. [32]	Vascular events (myocardial infarction, ischaemic stroke, coronary events)	MEDLINE, EMBASE, BIOSIS Previews, Cochrane CENTRAL, Conference Proceedings Citation Index Science, Google Scholar, ProQuest Dissertation Abstracts, Scopus, Science Citation Index Expanded	34	Myocardial infarction: RR 1.23 (95 % CI 1.15–1.31) Ischaemic stroke: RR 1.05 (95 % CI 1.01–1.09) Coronary events: RR 1.24 (1.10–1.39)

RR relative risk, *CI* cofidential interval

Table 6 Systematic reviews of shift work and gastrointestinal disorder

Reference	Target disease	Data base	Number of selected studies (number of studies indicating a significantly high risk)	Meta-analysis
Knutsson and Boggild [33]	Gastrointestinal disorders	MEDLINE	GI symptoms: 6(4) peptic ulcer disease: 6(5) functional GI disease: 3(2)	(−)

GI gastgro-intestinal

Table 7 Systematic reviews of types of shift work and health-related outcomes

References	Contents of review	Data base	Number of selected studies	Outcome of review
Bambra et al. [34]	The effects on health and work–life balance of organizational-level interventions that redesign shift work schedules.	Twenty-seven electronic databases	26	Beneficial effects on health and work–life balance: (1) switching from slow to fast rotation, (2) changing from backward to forward rotation, and (3) selfscheduling of shifts.
Vedaa et al. [35]	Quick returns (i.e. 11.0 h or less between two consecutive shifts) and outcome measures of health, sleep, functional ability and work-life balance	Web of Science, Pubmed, PsycINFO	22	There were some indications of detrimental effects of quick returns on proximate problems (e.g. sleep, sleepiness and fatigue)

promote atherogenesis and may then eventually increase the risk of onset of cardiovascular diseases [41, 43, 45]. In addition, immune responses may decrease the number of natural killer cells, which would lower immune defense and then increase the risk of onset of cancer [42, 52, 53]. Oxidative stress may promote atherogenesis by enhancing pro-oxidative status and reducing anti-oxidative status, thus increasing the risk of cancer onset [46, 54].

Conclusion

With regard to the effects of shift work on health, systematic review studies have been performed on associations with sleep, cancer, metabolic endocrine disorders, reproduction, cardiovascular diseases, and gastrointestinal disorders. There have also been systematic reviews on the associations between types of shift work and health-related

outcomes. Meta-analyses based on quantitative combination of the data from these studies showed that shift work significantly increased the risk of the following disorders: breast cancer, diabetes mellitus, preterm delivery, abortion, low birth weight, small-for-gestational-age infants, menstrual disruption, infertility, ischemic heart disease, and ischemic stroke. Some systematic studies indicated sufficient epidemiological evidence based on accumulated data from previous studies, including those obtained from meta-analyses. In contrast, not a few systematic review studies concluded that the evidence for effects on health was inconclusive due to the fact that the number of studies selected for such reviews had been insufficient. Similarly, it appears that insufficient epidemiological evidence has been accumulated for detection of differences in health effects based on shift work types. Therefore, further development of such studies is needed.

Acknowledgments This study was funded by Health Labor Sciences Research Grant (H25-JUNKANKITOU-IPPAN-007) from the Japanese Ministry of Health Labor and Welfare.

Compliance with ethical standards

Conflicts of interest The authors have no conflict of interest to declare.

References

1. IARC monographs on the evaluation of carcinogenic risks to humans. Painting, firefighting, and shiftwork. IARC monographs on the evaluation of carcinogenic risks to humans. vol 98. Lyon, France: International Agency for Research on Cancer, WHO Press; 2010, pp. 1017–1606. http://monographs.iarc.fr/ENG/Monographs/vol98/mono98.pdf.
2. McMenamin TM. Time to work: recent trends in shift work and flexible schedules, A. Mon Lab Rev. 2007;130:3.
3. Straif K, Baan R, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, Altieri A, Benbrahim-Tallaa L, Cogliano V. Carcinogenicity of shift-work, painting, and fire-fighting. Lancet Oncol. 2007;8:1065–6.
4. Statistics and Information Department of the Minister's Secretariat of the Ministry of Health. Survey on State of Employees' Health 1995–2014 (in Japanese). http://www.mhlw.go.jp/toukei/list/list46-50.html. Accessed 11 Nov 2015.
5. Vogel M, Braungardt T, Meyer W, Schneider W. The effects of shift work on physical and mental health. J Neural Transm. 2012;119:1121–32.
6. Niu SF, Chung MH, Chen CH, Hegney D, O'Brien A, Chou KR. The effect of shift rotation on employee cortisol profile, sleep quality, fatigue, and attention level: a systematic review. J Nurs Res. 2011;19:68–81.
7. Linton SJ, Kecklund G, Franklin KA, Leissner LC, Sivertsen B, Lindberg E, Svensson AC, Hansson SO, Sundin O, Hetta J, Bjorkelund C, Hall C. The effect of the work environment on future sleep disturbances: a systematic review. Sleep Med Rev. 2015;23:10–9.
8. Megdal SP, Kroenke CH, Laden F, Pukkala E, Schernhammer ES. Night work and breast cancer risk: a systematic review and meta-analysis. Eur J Cancer. 2005;41:2023–32.
9. Kolstad HA. Nightshift work and risk of breast cancer and other cancers—a critical review of the epidemiologic evidence. Scand J Work Environ Health. 2008;34:5–22.
10. Sigurdardottir LG, Valdimarsdottir UA, Fall K, Rider JR, Lockley SW, Schernhammer E, Mucci LA. Circadian disruption, sleep loss, and prostate cancer risk: a systematic review of epidemiologic studies. Cancer Epidemiol Biomark Prev. 2012;21:1002–11.
11. Ijaz S, Verbeek J, Seidler A, Lindbohm ML, Ojajarvi A, Orsini N, Costa G, Neuvonen K. Night-shift work and breast cancer—a systematic review and meta-analysis. Scand J Work Environ Health. 2013;39:431–47.
12. Jia Y, Lu Y, Wu K, Lin Q, Shen W, Zhu M, Huang S, Chen J. Does night work increase the risk of breast cancer? A systematic review and meta-analysis of epidemiological studies. Cancer Epidemiol. 2013;37:197–206.
13. Kamdar BB, Tergas AI, Mateen FJ, Bhayani NH, Oh J. Night-shift work and risk of breast cancer: a systematic review and meta-analysis. Breast Cancer Res Treat. 2013;138:291–301.
14. Wang F, Yeung KL, Chan WC, Kwok CC, Leung SL, Wu C, Chan EY, Yu IT, Yang XR, Tse LA. A meta-analysis on dose-response relationship between night shift work and the risk of breast cancer. Ann Oncol. 2013;24:2724–32.
15. van Drongelen A, Boot CR, Merkus SL, Smid T, van der Beek AJ. The effects of shift work on body weight change—a systematic review of longitudinal studies. Scand J Work Environ Health. 2011;37:263–75.
16. Canuto R, Garcez AS, Olinto MT. Metabolic syndrome and shift work: a systematic review. Sleep Med Rev. 2013;17:425–31.
17. Staufenbiel SM, Penninx BW, Spijker AT, Elzinga BM, van Rossum EF. Hair cortisol, stress exposure, and mental health in humans: a systematic review. Psychoneuroendocrinology. 2013;38:1220–35.
18. Manenschijn L, van Kruysbergen RG, de Jong FH, Koper JW, van Rossum EF. Shift work at young age is associated with elevated long-term cortisol levels and body mass index. J Clin Endocrinol Metab. 2011;96:E1862–5.
19. Knutsson A, Kempe A. Shift work and diabetes—a systematic review. Chronobiol Int. 2014;31:1146–51.
20. Gan Y, Yang C, Tong X, Sun H, Cong Y, Yin X, Li L, Cao S, Dong X, Gong Y, Shi O, Deng J, Bi H, Lu Z. Shift work and diabetes mellitus: a meta-analysis of observational studies. Occup Environ Med. 2015;72:72–8.
21. Ulhoa MA, Marqueze EC, Burgos LG, Moreno CR. Shift work and endocrine disorders. Int J Endocrinol. 2015;2015:826249.
22. Bonzini M, Coggon D, Palmer KT. Risk of prematurity, low birthweight and pre-eclampsia in relation to working hours and physical activities: a systematic review. Occup Environ Med. 2007;64:228–43.
23. Quansah R, Jaakkola JJ. Occupational exposures and adverse pregnancy outcomes among nurses: a systematic review and meta-analysis. J Womens Health (Larchmt). 2010;19:1851–62.
24. Bonzini M, Palmer KT, Coggon D, Carugno M, Cromi A, Ferrario MM. Shift work and pregnancy outcomes: a systematic review with meta-analysis of currently available epidemiological studies. BJOG. 2011;118:1429–37.
25. Bonde JP, Jorgensen KT, Bonzini M, Palmer KT. Miscarriage and occupational activity: a systematic review and meta-analysis regarding shift work, working hours, lifting, standing, and physical workload. Scand J Work Environ Health. 2013;39:325–34.
26. Chau YM, West S, Mapedzahama V. Night work and the reproductive health of women: an integrated literature review. J Midwifery Womens Health. 2014;59:113–26.
27. Stocker LJ, Macklon NS, Cheong YC, Bewley SJ. Influence of shift work on early reproductive outcomes: a systematic review and meta-analysis. Obstet Gynecol. 2014;124:99–110.

28. van Melick MJ, van Beukering MD, Mol BW, Frings-Dresen MH, Hulshof CT. Shift work, long working hours and preterm birth: a systematic review and meta-analysis. Int Arch Occup Environ Health. 2014;87:835–49.
29. Frost P, Kolstad HA, Bonde JP. Shift work and the risk of ischemic heart disease—a systematic review of the epidemiologic evidence. Scand J Work Environ Health. 2009;35:163–79.
30. Togo F, Takahashi M. Heart rate variability in occupational health—a systematic review. Ind Health. 2009;47:589–602.
31. Hwang WJ, Hong O. Work-related cardiovascular disease risk factors using a socioecological approach: implications for practice and research. Eur J Cardiovasc Nurs. 2012;11:114–26.
32. Vyas MV, Garg AX, Iansavichus AV, Costella J, Donner A, Laugsand LE, Janszky I, Mrkobrada M, Parraga G, Hackam DG. Shift work and vascular events: systematic review and meta-analysis. BMJ. 2012;345:e4800.
33. Knutsson A, Boggild H. Gastrointestinal disorders among shift workers. Scand J Work Environ Health. 2010;36:85–95.
34. Bambra CL, Whitehead MM, Sowden AJ, Akers J, Petticrew MP. Shifting schedules: the health effects of reorganizing shift work. Am J Prev Med. 2008;34:427–34.
35. Vedaa Ø, Harris A, Bjorvatn B, Waage S, Sivertsen B, Tucker P, Pallesen S. Systematic review of the relationship between quick returns in rotating shift work and health-related outcomes. Ergonomics. 2015;1–14. doi:10.1080/00140139.2015.1052020.
36. Faraut B, Bayon V, Leger D. Neuroendocrine, immune and oxidative stress in shift workers. Sleep Med Rev. 2013;17:433–44.
37. Saksvik IB, Bjorvatn B, Hetland H, Sandal GM, Pallesen S. Individual differences in tolerance to shift work—a systematic review. Sleep Med Rev. 2011;15:221–35.
38. Meerlo P, Sgoifo A, Suchecki D. Restricted and disrupted sleep: effects on autonomic function, neuroendocrine stress systems and stress responsivity. Sleep Med Rev. 2008;12:197–210.
39. Scheer FA, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. Proc Natl Acad Sci USA. 2009;106:4453–8.
40. Dinges DF, Douglas SD, Zaugg L, Campbell DE, McMann JM, Whitehouse WG, Orne EC, Kapoor SC, Icaza E, Orne MT. Leukocytosis and natural killer cell function parallel neurobehavioral fatigue induced by 64 h of sleep deprivation. J Clin Invest. 1994;93:1930–9.
41. Faraut B, Boudjeltia KZ, Dyzma M, Rousseau A, David E, Stenuit P, Franck T, Van Antwerpen P, Vanhaeverbeek M, Kerkhofs M. Benefits of napping and an extended duration of recovery sleep on alertness and immune cells after acute sleep restriction. Brain Behav Immun. 2011;25:16–24.
42. Logan RW, Zhang C, Murugan S, O'Connell S, Levitt D, Rosenwasser AM, Sarkar DK. Chronic shift-lag alters the circadian clock of NK cells and promotes lung cancer growth in rats. J Immunol. 2012;188:2583–91.
43. van Leeuwen WM, Lehto M, Karisola P, Lindholm H, Luukkonen R, Sallinen M, Harma M, Porkka-Heiskanen T, Alenius H. Sleep restriction increases the risk of developing cardiovascular diseases by augmenting proinflammatory responses through IL-17 and CRP. PLoS One. 2009;4:e4589.
44. Vgontzas AN, Pejovic S, Zoumakis E, Lin HM, Bixler EO, Basta M, Fang J, Sarrigiannidis A, Chrousos GP. Daytime napping after a night of sleep loss decreases sleepiness, improves performance, and causes beneficial changes in cortisol and interleukin-6 secretion. Am J Physiol Endocrinol Metab. 2007;292:E253–61.
45. Boudjeltia KZ, Faraut B, Esposito MJ, Stenuit P, Dyzma M, Antwerpen PV, Brohee D, Vanhamme L, Moguilevsky N, Vanhaeverbeek M, Kerkhofs M. Temporal dissociation between myeloperoxidase (MPO)-modified LDL and MPO elevations during chronic sleep restriction and recovery in healthy young men. PLoS One. 2011;6:e28230.
46. Everson CA, Laatsch CD, Hogg N. Antioxidant defense responses to sleep loss and sleep recovery. Am J Physiol Regul Integr Comp Physiol. 2005;288:R374–83.
47. Vinogradova IA, Anisimov VN, Bukalev AV, Ilyukha VA, Khizhkin EA, Lotosh TA, Semenchenko AV, Zabezhinski MA. Circadian disruption induced by light-at-night accelerates aging and promotes tumorigenesis in young but not in old rats. Aging (Albany NY). 2010;2:82–92.
48. Knutsson A. Health disorders of shift workers. Occup Med (Lond). 2003;53:103–8.
49. Lo SH, Lin LY, Hwang JS, Chang YY, Liau CS, Wang JD. Working the night shift causes increased vascular stress and delayed recovery in young women. Chronobiol Int. 2010;27:1454–68.
50. Puttonen S, Kivimaki M, Elovainio M, Pulkki-Raback L, Hintsanen M, Vahtera J, Telama R, Juonala M, Viikari JS, Raitakari OT, Keltikangas-Jarvinen L. Shift work in young adults and carotid artery intima-media thickness: the cardiovascular risk in young finns study. Atherosclerosis. 2009;205:608–13.
51. Wang XS, Armstrong ME, Cairns BJ, Key TJ, Travis RC. Shift work and chronic disease: the epidemiological evidence. Occup Med (Lond). 2011;61:78–89.
52. Gungor N, Godschalk RW, Pachen DM, Van Schooten FJ, Knaapen AM. Activated neutrophils inhibit nucleotide excision repair in human pulmonary epithelial cells: role of myeloperoxidase. Faseb J. 2007;21:2359–67.
53. Imai K, Matsuyama S, Miyake S, Suga K, Nakachi K. Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. Lancet. 2000;356:1795–9.
54. Poeggeler B, Thuermann S, Dose A, Schoenke M, Burkhardt S, Hardeland R. Melatonin's unique radical scavenging properties—roles of its functional substituents as revealed by a comparison with its structural analogs. J Pineal Res. 2002;33:20–30.

Sleep Medicine 25 (2016) 130–138

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Nationwide epidemiological study of insomnia in Japan

Osamu Itani [a], Yoshitaka Kaneita [a,*], Takeshi Munezawa [b], Kazuo Mishima [c], Maki Jike [d], Sachi Nakagome [d], Mikiko Tokiya [a], Takashi Ohida [d]

[a] Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, Oita, Japan
[b] Socio-Tech Institute Co., Ltd., Tokyo, Japan
[c] National Center of Neurology and Psychiatry, Tokyo, Japan
[d] Division of Public Health, Department of Social Medicine, Nihon University School of Medicine, Tokyo, Japan

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form 11 May 2016
Accepted 15 May 2016
Available online 30 August 2016

Keywords:
Prevalence
Associated factors
Season
Region
Winter
Summer

ABSTRACT

Background: This study was a nationwide epidemiological study of insomnia in Japan. It was conducted because very few studies on this topic have previously been performed for the general Japanese population.
Methods: An interview survey on symptoms of insomnia (difficulty initiating sleep, difficulty maintaining sleep with difficulty resuming sleep, and early morning awakening with difficulty resuming sleep) and daytime dysfunction was conducted on the general nationwide population in the winter (February) and summer (August) of 2008. Data from 2614 participants who provided valid responses (age range 20–95 years, valid response rate 54.2%) were analyzed.
Results: The prevalence of difficulty initiating sleep, difficulty maintaining sleep with difficulty resuming sleep, and early morning awakening with difficulty resuming sleep was 8.3%, 5.8%, and 5.8%, respectively, in men, and 11.0%, 8.1%, and 7.4%, respectively, in women. The prevalence of insomnia was 12.2% in men and 14.6% in women, and the prevalence of insomnia with daytime dysfunction was 3.2% in men and 4.2% in women. The results of logistic regression analyses indicated that the factors aggravating insomnia for men were unemployment and having mental health issues, and for women they were being aged ≥70 years, completing fewer years of schooling, and having mental health issues. Seasonality and regionality in association with insomnia were also examined, but no significant associations were found.
Conclusion: In the present survey, insomnia was defined by using criteria that were closer to the clinical diagnostic criteria (eg, coexistence of both difficulty resuming sleep and daytime dysfunction was considered). Therefore, it is believed that the results of this study were representative of the clinical actuality of insomnia in Japan.

1. Introduction

Previous studies have found that insomnia is a common sleep disorder and an important risk factor for various physical and mental disorders [1]. High-quality epidemiological studies are extremely important for accurately understanding the actual status of insomnia and taking measures to improve this condition. As such, studies on the prevalence of, and factors associated with, insomnia have been actively conducted in many areas of the world [2].

In 1997, Doi et al. conducted a nationwide epidemiological study of insomnia in Japan; they used 2800 individuals aged ≥20 years who were randomly selected from the general population [3]. This self-administered questionnaire survey defined insomnia as the presence of any of the following three symptoms: difficulty initiating sleep (DIS), difficulty maintaining sleep (DMS), and early morning awakening (EMA). The results indicated that the prevalence of insomnia was 17.3% in men and 21.5% in women. Another nationwide study was conducted in 1997 by Kim et al. on 3030 individuals aged ≥20 years who were randomly selected from the general population [4]. The prevalence of insomnia determined from this interview survey was 20.5% in men and 22.3% in women. These are the only two nationwide-scale epidemiological studies of insomnia that have been performed in Japan.

Therefore, the present study was designed to address the shortcomings of the previous studies in order to clarify the actual status of insomnia in Japan. The first consideration was the definition of insomnia used to evaluate its prevalence. Differences in the definition of insomnia and the use of various criteria to define it have resulted in substantially large differences in the reported prevalence of insomnia in previous epidemiological studies worldwide [2].

In the two previous nationwide epidemiological studies of insomnia in Japan, insomnia was defined as merely having symptoms of insomnia [3,4]. In those studies, the questions regarding DMS and

* Corresponding author. Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, Hasama-machi, Yufu-shi, Oita 879-5593, Japan. Fax: +81 97 586 5739.
E-mail address: nusmpublichealth@gmail.com (Y. Kaneita).

http://dx.doi.org/10.1016/j.sleep.2016.05.013
1389-9457/

EMA were inadequate because they did not contain the element of difficulty resuming sleep (DRS). Therefore, the previous studies merely queried the presence of awakening from sleep. Several previous studies have indicated that mere nocturnal awakening is not always a pathological condition [5,6]. Ohayon et al. proposed that nighttime awakening must be assessed in terms of both the frequency and coexistence of DRS [5]. Based on the above findings, the present study considered the coexistence of DRS when investigating DMS and EMA as symptoms of insomnia.

Previous epidemiological studies of insomnia have reported that individuals who had daytime dysfunction (such as sleepiness, fatigue, mood disturbance, cognitive difficulties, social impairment, or occupational impairment) [7] in addition to insomnia required more full-scale health care, and were more likely to have mental disorders or organic diseases than those who had insomnia alone [2,8]. Therefore, in the present study, the presence/absence of daytime dysfunctions was also considered in the definition of insomnia. It was expected that considering the coexistence of daytime dysfunction would increase the clinical utility of any epidemiological study of insomnia.

Furthermore, while the two previous Japanese studies did not examine varying prevalences of insomnia based on differences in seasons or regions, the present study examined the seasonality and regionality of insomnia. To this end, the survey was conducted twice – in the summer and winter – and information was collected regarding the residence (place, size of the city, etc.) in each survey.

The present study had two objectives. First, it aimed to perform an epidemiological study of insomnia with high clinical utility by using new defining criteria such as the coexistence of DRS with symptoms of insomnia, and the coexistence of daytime dysfunction with insomnia. Second, it aimed to examine the seasonality or regionality of insomnia in Japan.

2. Materials and methods

2.1. Participants

The participants were sampled using the stratified three-stage random sampling method. The data were obtained by first performing stratification, in which: (1) the municipalities across the country were classified into 12 blocks using prefecture or city as a unit, and (2) each block was further divided into 19 large cities, other cities, and towns/villages according to their size. Based on a population aged ≥20 years, a total of 4000 samples were proportionally allocated in strata of the size of blocks and cities. In the three-stage extraction (to match the population distribution of Japan), predetermined criteria were used to determine the number of extractions for:

(1) Survey locations: the number of samples were proportionally allocated according to the estimated population of each stratum, and then adjusted so that the number of samples in one survey area became approximately 25 as a standard. Through this method, the number of survey areas (primary sampling units) was determined to be 157.
(2) Households: for selection of households (secondary sampling units), using a residential map database, every fourth house in the survey areas selected, as described in (1), was chosen as a subject household.
(3) Participants: upon contact with any member of a subject household, the researcher asked for the member of the subject household who would first reach his/her 20th birthday or older after the first of the survey month, and that person was selected as the subject. If the subject was not at home, the researcher was instructed to visit the house at least three times. If the researcher was unable to meet the subject, the interview with that subject was determined to be impossible. Other indices were extracted randomly. This tried to ensure an authentic representation of the general Japanese population.

The survey was conducted in February 2008 (winter survey) and August 2008 (summer survey). In the winter survey, the researchers were able to make contact with a member from 2449/4000 subject households; 1308 who agreed to participate in the survey were interviewed. Trained interviewers visited the selected households and interviewed persons aged ≥20 years. Verbal informed consent was obtained from the participants, and their privacy was protected in accordance with the Declaration of Helsinki. Stratified random sampling was similarly performed for the summer survey, and 1306/2371 selected subjects agreed to participate in the interview survey.

2.2. Interview method

The present survey employed a face-to-face interview method. Well-trained researchers visited the subject households and interviewed the subjects using a questionnaire that had been prepared beforehand. A structured interview method was employed, wherein a researcher read the questions exactly as they appeared on the questionnaire. Sheets indicating the answer options were presented to the subjects for their selection of a response.

2.3. Measures

The questionnaire inquired about the following four items, which are described in detail below: (1) basic attributes, (2) questions regarding the symptoms of insomnia, (3) questions regarding daytime dysfunction, and (4) mental health status.

(1) Basic attributes
  (a) Gender and age
  (b) Occupation: Agricultural, forestry, or fishery worker/commercial, industrial, or service business worker/clerk/laborer/freelancer or managerial staff/housewife without a job/student/others or unemployed.

State of occupation: For the statistical analyses, agricultural, forestry, or fishery workers, commercial, industrial, or service business workers, clerks, laborers, freelancers, and managerial staff were categorized as "with occupation." Housewives without a job were categorized as "housewives." Students, others, or the unemployed were categorized as "without occupation."

  (c) Regions: Hokkaido and Tohoku/Kanto, Keihin, and Koshinetsu/Hokuriku and Tokai/Kinki and Hanshin/Chugoku and Shikoku/Kyushu.
  (d) Size of the city of residence: 18 large cities; 17 government-decreed cities, and Tokyo Metropolis)/city/municipality.
  (e) Years of schooling completed: junior high school/senior high school/college or university.

(2) Questions regarding the symptoms of insomnia

The following three questions about symptoms of insomnia were asked:

  (a) Question related to DIS

Q: "Have you had difficulty initiating sleep over the past one month?"

A: Never/seldom/sometimes/often/always

  (b) Question related to DMS with DRS

Q: "Have you awoken at night and had difficulty falling asleep again over the past one month?"

A: Never/seldom/sometimes/often/always

  (c) Question related to EMA with DRS

Q: "Have you awoken early in the morning (at dawn) and had difficulty falling asleep again over the past one month?"

A: Never/seldom/sometimes/often/always

Participants who selected often or always for questions (a), (b), and (c) were defined as having DIS, DMS with DRS, and EMA with DRS, respectively. Analysis of these data was subsequently performed. In addition, participants who had at least one of DIS, DMS with DRS, or EMA with DRS, were defined as having insomnia, and analysis of these data was performed.

(3) Questions regarding daytime dysfunction

The following two questions about daytime dysfunction were asked:

(a) Q: "Have you been excessively sleepy during the daytime and unable to prevent yourself from falling asleep when you must not sleep over the past one month?"

A: Never/seldom/sometimes/often/always

(b) Q: "Have you felt physically or mentally disordered or had any trouble in your daily life as a result of lack of good sleep over the past one month?"

A: Never/seldom/sometimes/often/always

Participants who selected often or always for question (a) or (b) were defined as having daytime dysfunction, and analysis of this data was performed. In addition, participants who had both insomnia and daytime dysfunction were defined as having insomnia with daytime dysfunction, and analysis of this data was performed.

(4) Mental health status

The following two questions about mental health status were asked:

(a) Q: "Have you felt an unusual amount of unhappiness and depression over the past one month?"

A: Not at all/no more than usual/more than usual/much more than usual

(b) Q: "Have you been able to enjoy normal activities more than usual over the past one month?"

A: More so than usual/same as usual/less than usual/much less than usual

To evaluate the mental health statuses of the respondents, two independent factors (depression/anxiety and decrease in positive feeling) included in the 12-item General Health Questionnaire (GHQ-12) [9,10] were used, and one item from each factor was selected for the total score. One of the items from the depression/anxiety factor was evaluated. One of the items from the decrease in positive feeling factor was also evaluated. Each item described a symptom, and there were four possible answers: the two answers that indicated a lack of the symptom were assigned a rating of 0; the two answers that indicated a presence of the symptom were assigned a rating of 1. Thus, the overall score fell within the range of 0–2, and, accordingly, the higher the total score, the poorer the state of mental health. In the present study, participants who had total scores of ≥1 were considered to have poor mental health. Previous studies have shown that evaluation of mental health status using depression symptoms with the GHQ-12 and with this cut-off point has a sensitivity of 87.0% and a specificity of 85.1% [11]. These methods and cut-off were employed in other large-scale epidemiological studies [12–20].

2.4. Statistical analysis

First, the participants were stratified separately according to sex. The prevalence of DIS, DMS with DRS, EMA with DRS, insomnia, and insomnia with daytime dysfunction according to age class, state of occupation, region, size of the city of residence, years of schooling completed, season (February or August) of the survey, and mental health status were calculated. The Chi-squared test was used to determine statistical significance.

Second, multiple logistic regression analyses based on sex were performed by using DIS, DMS with DRS, EMA with DRS, insomnia, and insomnia with daytime dysfunction as the response variables. As the predictor variables, age class, state of occupation, region, size of the city of residence, years of schooling completed, season (February or August) of the survey, and mental health status were used. Significance was set at $p < 0.05$. All analyses were performed using IBM SPSS Statistics, version 22 J for Windows (IBM Corp., Somers, NY, USA).

2.5. Ethical considerations

The following ethical considerations were taken into account: (1) participation was voluntary, but informed consent was required from the participant; (2) researchers who performed data collection were different from those who performed statistical analysis, to prevent access to the participants' personal data; (3) collected data were coded to protect personal information and maintain confidentiality; and (4) prior approval was obtained from the Ethics Committee for Epidemiological Study of Nihon University School of Medicine.

3. Results

From the total of 4820 subjects selected for the two surveys, 2614 (men 1189, women 1425) responded to the questions (response rate 54.2%). The ages of the participants ranged from 20 to 95 years. The mean age (±SD) of the participants in total was 52.4 years (±16.9); the mean ages for men and women were 52.2 years (±17.7) and 52.6 years (±16.5), respectively. The backgrounds of the participants based on sex are shown in Table 1.

Of the 2614 respondents, nine failed to complete data about insomnia symptoms. The prevalence of insomnia symptoms (DIS, DMS with DRS, and EMA with DRS) based on sex is shown in Table 2. The prevalence of each symptom was as follows: DIS, 8.3% in men and 11.0% in women; DMS with DRS, 5.8% in men and 11.0% in women; and EMA with DRS, 5.8% in men and 7.4% in women. The Chi-squared test indicated that DIS was significantly associated with mental health status ($p < 0.01$) for men and with age class ($p < 0.01$), state of occupation ($p = 0.02$), years of schooling completed ($p < 0.01$), and mental health status ($p < 0.01$) for women. DMS with DRS was significantly associated with state of occupation ($p = 0.01$) and mental health status ($p < 0.01$) for men and with age class ($p < 0.01$), state of occupation ($p < 0.01$), years of schooling completed ($p = 0.03$), and mental health status ($p < 0.01$) for women. EMA with DRS was significantly associated with state of occupation ($p < 0.01$) and mental health status ($p < 0.01$) for men and with age class ($p < 0.01$), state of occupation ($p = 0.02$), and mental health status ($p < 0.01$) for women. With regard to the season of the survey, region, and size of the city of residence, no significant associations with DIS, DMS with DRS, or EMA with DRS were found.

The prevalence of insomnia and insomnia with daytime dysfunction based on sex is shown in Table 3. The prevalence of insomnia was 12.2% in men and 14.6% in women, and the prevalence of insomnia with daytime dysfunction was 3.2% in men and 4.2% in women. The Chi-squared test indicated that insomnia was significantly associated with mental health status ($p < 0.01$) for men and with age class ($p < 0.01$), state of occupation ($p = 0.01$), years of schooling completed ($p < 0.01$), and mental health status ($p < 0.01$) for women. Insomnia with daytime dysfunction was significantly associated with size of the city ($p = 0.02$) and mental health status ($p < 0.01$) for men and with age class ($p < 0.01$), state of occupation ($p < 0.01$), years of schooling completed ($p = 0.05$), and mental health status ($p < 0.01$) for women. With regard to the season of the survey and region, no significant associations were found with either insomnia or insomnia with daytime dysfunction.

Table 1
Background of the survey subjects.

	Male		Female	
	N	%	N	%
Age class				
20–29 years	142	11.9	146	10.2
30–39 years	214	18.0	226	15.9
40–49 years	181	15.2	241	16.9
50–59 years	195	16.4	283	19.9
60–69 years	224	18.8	277	19.4
≥70 years	233	19.6	252	17.7
Occupation				
Agricultural, forestry, or fishery worker	27	2.3	26	1.8
Commercial, industrial, or service business worker	162	13.6	109	7.6
Clerk	269	22.6	201	14.1
Laborer	295	24.8	242	17.0
Freelancer or managerial staff	64	5.4	16	1.1
Housewife without a job	0	0.0	677	47.5
Student	31	2.6	19	1.3
Others or unemployed	341	28.7	135	9.5
Years of schooling completed				
Junior high school	149	12.6	215	15.2
Senior high school	599	50.5	782	55.1
College or university	439	37.0	421	29.7
Seasons of the survey				
Winter (February)	589	49.5	719	50.5
Summer (August)	600	50.5	706	49.5
Regions				
Hokkaido and Tohoku	155	13.0	174	12.2
Kanto, Keihin, and Koshinetsu	433	36.4	483	33.9
Hokuriku and Tokai	182	15.3	201	14.1
Kinki and Hanshin	166	14.0	248	17.4
Chugoku and Shikoku	112	9.4	153	10.7
Kyushu	141	11.9	166	11.6
Size of the city of residence				
18 large cities	284	23.9	357	25.1
City	783	65.9	910	63.9
Municipality	122	10.3	158	11.1
Mental health status				
Good	872	73.3	969	68.0
Poor	317	26.7	456	32.0

Subject with missing data was excluded from the analysis.

The results of multiple logistic regression analyses for separately stratified men and women by using DIS, DMS with DRS, or EMA with DRS as response variables are shown in Table 4. The adjusted odds ratios (AORs) for DIS, DMS with DRS, and EMA with DRS were significantly increased for mentally unhealthy subjects, both in men and women, when those in good mental health were used as the reference. In addition, for women, the AOR of DIS was significantly decreased for college or university graduates when junior or senior high school graduates were used as the reference.

The results of multiple logistic regression analyses for separately stratified men and women by using insomnia or insomnia with daytime dysfunction as a response variable are shown in Table 5. The AORs for insomnia and insomnia with daytime dysfunction were significantly increased for mentally unhealthy male and female subjects when those in good mental health were used as the reference. For men, the AORs for insomnia and insomnia with daytime dysfunction were significantly increased for those without occupation when those with occupation were used as the reference. In women, the AOR for insomnia was significantly increased for those aged ≥70 years when those aged 20–29 years were used as the reference. In addition, the AOR for college or university graduates was significantly decreased in women when junior or senior high school graduates were used as the reference. Moreover, in men, the AOR for insomnia with daytime dysfunction was significantly increased for those living in municipalities when those living in cities were used as the reference.

4. Discussion

4.1. Prevalence of insomnia symptoms or insomnia

The overall prevalence of insomnia in the present study was 13.5%, including DIS (9.8%), DMS with DRS (7.1%), and EMA with DRS (6.7%). The prevalence of insomnia, DMS, and EMA were substantially lower in the present study than in previous studies: insomnia 17.3–21.4% [3,4]; DIS 8.3–12.6% [3,4]; DMS 12.9–16.2% [3,4]; EMA 8.0% [4]. One reason for these differences may have been the stricter definition of insomnia, which was closer to the clinical definition (ie, considering the coexistence of DRS in DMS and EMA). In previous studies conducted in Japan, the prevalence of insomnia might have been overestimated because individuals who had DMS or EMA, but could not be defined as having insomnia, were defined as having insomnia. Therefore, the results of the present study may provide a more clinically accurate representation of insomnia in Japan.

In this study, the overall prevalence of insomnia with daytime dysfunction was 3.8%. This was the first study that considered the coexistence of daytime dysfunction, as this was not taken into account in previous epidemiological studies of insomnia in Japan. Ohayon classified previous epidemiological studies of insomnia into four categories according to the definitions employed, and reported the prevalence of insomnia in each category. The prevalence of insomnia ranged from 4 to 6% when the strictest definition of insomnia was used (ie, employing diagnostic criteria (such as those provided in the DSM-III-R [21] and DSM-IV [7] and the International Classification of Sleep Disorders [22]) [2]. In the present study, evaluation criteria that were similar to the clinical diagnostic criteria for insomnia were adopted. Specifically, the duration of insomnia symptoms was set to one month and the coexistence of daytime dysfunctions was considered in the survey questions. Therefore, the study results may be more clinically accurate.

The prevalence rates of the symptoms of insomnia, insomnia itself, and insomnia with daytime dysfunction were higher in women than in men. Lichstein et al. reviewed 33 previous studies that compared the prevalence of insomnia based on sex and reported that the prevalence of insomnia symptoms, and of insomnia itself, was higher in women than in men [23]. The results of the present study are consistent with this finding.

4.2. Associated factors of insomnia

The results of multivariate analyses indicated that the factors aggravating insomnia for men were being without occupation and having mental health issues, and for women they were being ≥70 years (compared with the 20–29 year age group), completing fewer years of schooling, and having mental health issues. In two previous epidemiological studies of insomnia in Japan [3,4], the factors associated with insomnia for men were never having been married, being widowed, or unemployed; for women they were older age, being unemployed, lack of habitual exercise, poor perceived health, psychological stress, and being unable to cope with stress.

Lichstein et al. reviewed 20 studies and reported that the prevalence and severity of insomnia were associated with age in 60% of those studies [23]. These authors found strong evidence for increased DMS with age, but modest evidence for increased DIS and EMA with age. Ohayon et al. performed a meta-analysis of sleep parameters reported in 65 studies of subjects ranging in age from five to 102 years [24]. They reported that sleep latency and the percentages of stage 1 and stage 2 sleep appeared to increase significantly with age, while the percentage of REM sleep decreased in adults. This indicates that, in the elderly, the mechanism for maintaining sleep is significantly impaired compared with the mechanism for initiating sleep. This result did not contradict the findings of the review by Lichstein et al. However, several other

Table 2
Prevalence of insomnia symptoms.

	DIS						DMS with DRS						EMA with DRS					
	Male			Female			Male			Female			Male			Female		
	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*
Overall	8.3	6.7–9.9		11.0	9.4–12.6		5.8	4.5–7.1		8.1	6.7–9.5		5.8	4.5–7.1		7.4	6.0–8.8	
Age class			0.57			<0.01			0.14			<0.01			0.36			<0.01
20–29 years	9.2	4.4–14.0		8.2	3.7–12.7		4.9	1.3–8.5		5.5	1.8–9.2		4.9	1.3–8.5		3.4	0.5–6.3	
30–39 years	6.5	3.2–9.8		8.9	5.2–12.6		5.1	2.2–8.0		4.0	1.4–6.6		3.7	1.2–6.2		4.5	1.8–7.2	
40–49 years	7.7	3.8–11.6		6.6	3.5–9.7		3.3	0.7–5.9		5.4	2.5–8.3		4.4	1.4–7.4		5.0	2.2–7.8	
50–59 years	11.4	6.9–15.9		12.8	8.9–16.7		4.7	1.7–7.7		8.8	5.5–12.1		5.7	2.4–9.0		8.5	5.3–11.7	
60–69 years	8.5	4.8–12.2		10.5	6.9–14.1		9.4	5.6–13.2		10.1	6.6–13.6		7.6	4.1–11.1		8.3	5.0–11.6	
≥70 years	7.3	4.0–10.6		17.5	12.8–22.2		6.5	3.3–9.7		12.7	8.6–16.8		7.8	4.3–11.3		12.3	8.2–16.4	
State of occupation			0.36			0.02			0.01			<0.01			<0.01			0.02
With occupation	7.9	6.0–9.8		8.6	6.3–10.9		4.7	3.2–6.2		5.6	3.7–7.5		4.5	3.1–5.9		5.1	3.3–6.9	
Housewife	0.0	0.0–0.0		12.0	9.6–14.4		0.0	0.0–0.0		9.3	7.1–11.5		0.0	0.0–0.0		8.9	6.8–11.0	
Without occupation	9.4	6.4–12.4		16.2	10.4–22.0		8.4	5.6–11.2		12.3	7.1–17.5		8.6	5.7–11.5		9.7	5.0–14.4	
Years of schooling completed			0.25			<0.01			0.42			0.03			0.07			0.07
Junior or senior high school	9.0	6.9–11.1		12.9	10.8–15.0		6.2	4.5–7.9		9.1	7.3–10.9		6.7	4.9–8.5		8.2	6.5–9.9	
College or university	7.1	4.7–9.5		6.9	4.5–9.3		5.0	3.0–7.0		5.7	3.5–7.9		4.1	2.2–6.0		5.5	3.3–7.7	
Seasons of the survey			0.82			0.62			0.23			0.34			0.64			0.07
Winter (February)	8.2	6.0–10.4		11.5	9.2–13.8		6.6	4.6–8.6		7.4	5.5–9.3		6.1	4.2–8.0		6.1	4.3–7.9	
Summer (August)	8.5	6.3–10.7		10.6	8.3–12.9		5.0	3.3–6.7		8.8	6.7–10.9		5.5	3.7–7.3		8.6	6.5–10.7	
Regions			0.50			0.27			0.59			0.92			0.88			0.62
Hokkaido and Tohoku	7.7	3.5–11.9		13.8	8.7–18.9		5.8	2.1–9.5		9.8	5.4–14.2		5.2	1.7–8.7		6.9	3.1–10.7	
Kanto, Keihin, and Koshinetsu	7.4	4.9–9.9		9.7	7.1–12.3		6.7	4.3–9.1		7.5	5.2–9.8		6.3	4.0–8.6		6.6	4.4–8.8	
Hokuriku and Tokai	10.0	5.6–14.4		11.5	7.1–15.9		6.6	3.0–10.2		9.0	5.0–13.0		5.5	2.2–8.8		7.0	3.5–10.5	
Kinki and Hanshin	9.0	4.6–13.4		13.7	9.4–18.0		4.2	1.1–7.3		7.3	4.1–10.5		6.0	2.4–9.6		7.3	4.1–10.5	
Chugoku and Shikoku	5.4	1.2–9.6		11.2	6.2–16.2		2.7	−0.3–5.7		7.8	3.6–12.0		3.6	0.1–7.1		7.2	3.1–11.3	
Kyushu	11.4	6.1–16.7		7.3	3.3–11.3		6.4	2.3–10.5		8.5	4.2–12.8		7.1	2.8–11.4		10.9	6.1–15.7	
Size of the city of residence			0.10			0.34			0.62			0.46			0.28			0.69
18 large cities	5.6	2.9–8.3		11.2	7.9–14.5		6.4	3.5–9.3		7.8	5.0–10.6		5.7	3.0–8.4		7.6	4.8–10.4	
City	8.8	6.8–10.8		11.6	9.5–13.7		5.4	3.8–7.0		8.6	6.8–10.4		5.4	3.8–7.0		7.6	5.9–9.3	
Municipality	11.5	5.8–17.2		7.6	3.5–11.7		7.4	2.8–12.0		5.7	2.1–9.3		9.0	3.9–14.1		5.7	2.1–9.3	
Mental health status			<0.01			<0.01			<0.01			<0.01			<0.01			<0.01
Good	5.5	4.0–7.0		7.0	5.4–8.6		4.3	2.9–5.7		4.7	3.4–6.0		4.3	2.9–5.7		4.8	3.4–6.2	
Poor	16.2	12.1–20.3		19.9	13.6–26.2		10.2	6.9–13.5		15.5	12.2–18.8		10.2	6.9–13.5		13.0	9.9–16.1	

Subject with missing data was excluded from the analysis.
p was calculated by χ^2 test.
Abbreviations: DIS, difficulty initiating sleep; DMS, difficulty maintaining sleep; DRS, difficulty resuming sleep; EMA, early morning awakening; CI, confidential interval.

Table 3
Prevalence of insomnia.

	Insomnia						Insomnia with daytime dysfunction					
	Male			Female			Male			Female		
	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*	%	95% CI	*p*
Overall	12.2	10.3–14.1		14.6	12.8–16.4		3.2	2.2–4.2		4.2	3.2–5.2	
Age class			0.84			<0.01			0.90			<0.01
20–29 years	9.9	5.0–14.8		10.3	5.4–15.2		2.9	0.1–5.7		3.4	0.5–6.3	
30–39 years	11.7	7.4–16.0		10.3	6.3–14.3		2.8	0.6–5.0		1.8	0.1–3.5	
40–49 years	11.0	6.4–15.6		9.1	5.5–12.7		2.8	0.4–5.2		2.9	0.8–5.0	
50–59 years	14.5	9.5–19.5		16.0	11.7–20.3		4.7	1.7–7.7		3.2	1.1–5.3	
60–69 years	12.9	8.5–17.3		15.9	11.6–20.2		3.1	0.8–5.4		4.3	1.9–6.7	
≥70 years	12.1	7.9–16.3		23.4	18.2–28.6		3.1	0.9–5.3		9.1	5.5–12.7	
State of occupation			0.05			0.01			0.06			<0.01
With occupation	10.9	8.8–13.0		11.5	8.9–14.1		2.6	1.5–3.7		2.5	1.2–3.8	
Housewife	0.0	0.0–0.0		16.4	13.6–19.2		0.0	0.0–0.0		4.9	3.3–6.5	
Without occupation	14.9	11.3–18.5		18.8	12.6–25.0		4.6	2.5–6.7		7.8	3.6–12.0	
Years of schooling completed			0.15			<0.01			0.11			0.05
Junior or senior high school	13.2	10.8–15.6		16.7	14.4–19.0		3.8	2.4–5.2		4.9	3.6–6.2	
College or university	10.3	7.5–13.1		10.0	7.1–12.9		2.1	0.8–3.4		2.6	1.1–4.1	
Seasons of the survey			0.82			0.32			0.95			0.83
Winter (February)	11.9	9.3–14.5		13.7	11.2–16.2		3.3	1.9–4.7		4.3	2.8–5.8	
Summer (August)	12.4	9.8–15.0		15.6	12.9–18.3		3.2	1.8–4.6		4.1	2.6–5.6	
Regions			0.14			0.78			0.58			0.47
Hokkaido and Tohoku	10.4	5.6–15.2		17.8	12.1–23.5		2.6	0.1–5.1		4.6	1.5–7.7	
Kanto, Keihin, and Koshinetsu	11.3	8.3–14.3		14.1	11.0–17.2		3.5	1.8–5.2		3.7	2.0–5.4	
Hokuriku and Tokai	15.0	9.8–20.2		14.6	9.7–19.5		5.1	1.9–8.3		3.5	0.9–6.1	
Kinki and Hanshin	13.3	8.1–18.5		15.7	11.2–20.2		3.0	0.4–5.6		4.4	1.8–7.0	
Chugoku and Shikoku	6.3	1.8–10.8		13.2	7.8–18.6		2.7	−0.3–5.7		7.2	3.1–11.3	
Kyushu	16.4	10.3–22.5		12.7	7.6–17.8		1.4	−0.5–3.3		3.0	0.4–5.6	
Size of the city of residence			0.12			0.74			0.02			0.25
18 large cities	9.9	6.4–13.4		15.2	11.5–18.9		2.5	0.7–4.3		4.3	2.2–6.4	
City	12.2	9.9–14.5		14.8	12.5–17.1		2.8	1.6–4.0		3.2	2.1–4.3	
Municipality	17.2	10.5–23.9		12.7	7.5–17.9		7.4	2.8–12.0		5.1	1.7–8.5	
Mental health status			<0.01			<0.01			<0.01			<0.01
Good	8.6	6.7–10.5		10.1	8.2–12.0		1.6	0.8–2.4		1.7	0.9–2.5	
Poor	22.3	17.7–26.9		24.6	20.6–28.6		7.7	4.7–10.7		9.7	7.0–12.4	

Subject with missing data was excluded from the analysis.
p was calculated by χ^2 test.
Abbreviations: CI, confidential interval.

epidemiological studies of insomnia have refuted any significant associations between age and insomnia after adjustment for other important comorbidities [25–27]. No definitive epidemiological conclusion regarding the association between age and insomnia has yet been drawn. In the present study, a significant association was only observed between insomnia and specific age classes (20–29 years and ≥70 years old) in women. However, no significant associations were observed between age and the symptoms of insomnia in men and women.

Some previous studies have reported associations between insomnia and education as well as socioeconomic status (income and employment). An epidemiological study of 3684 public servants working for local authorities in Japan reported a significant association between employment grade and the Pittsburgh Sleep Quality Index (PSQI) score (an index of insomnia) in men, but not in women [28]. A study of 94 women aged 61–90 years reported a significant association between PSQI score and socioeconomic status defined according to education and income [29]. In addition, Mezick et al. investigated the associations between socioeconomic status and subjective sleep quality defined by the PSQI score, and concluded that lower socioeconomic status was associated with lower subjective sleep quality [30].

Previous studies have reported associations between psychological disorders and insomnia. In particular, insomnia was the strongest risk factor for depression. According to the results of a systematic review of longitudinal epidemiological studies on associations between insomnia and onset of depression by Baglioni et al., in which 21 studies published between 1980 and 2010 were included in a meta-analysis, the overall OR for the prediction of depression in insomnia cases was 2.60 (95% CI 1.98–3.42) [31]. In addition, aggravation of sleep before onset of MDD was indicated [32]. Furthermore, chronic insomnia was shown to be a risk factor for reducing the effects of therapy for depression [33,34].

The present study observed sex-based differences in the prevalence of insomnia and factors associated with insomnia. Zhang et al. performed a meta-analysis of 29 papers and reported that the risk ratio (95% CI) for insomnia in women was significantly higher [1.41 (1.28–1.55)] than that in men [35]. They also performed a meta-analysis based on the subtypes of insomnia (DIS, DMS, EMA, and non-restorative sleep), and reported that all subtypes of insomnia had a significant female preponderance, with the exception of non-restorative sleep [35]. These findings are consistent with the results of the present study. The difference in the rise and fall of sleep-related hormone levels between men and women provides a physiological explanation for sex-based differences in insomnia [36].

4.3. Seasonality and regionality of insomnia in Japan

The present study also investigated the seasonality and regionality of insomnia in Japan. With regard to the seasonality of insomnia, the main focus has been on associations between the midnight sun in Polar Regions and sleep, and previous epidemiological studies have mainly been conducted in Northern Europe. These studies indicated that symptoms of insomnia are more common in summer and winter than in spring and autumn [37]. When summer and

Table 4
Factors associated with insomnia symptoms.

	DIS						DMS with DRS						EMA with DRS					
	Male			Female			Male			Female			Male			Female		
	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*
Age class			0.69			0.08			0.37			0.17			0.98			0.11
20–29 years	1.00			1.00			1.00			1.00			1.00			1.00		
30–39 years	0.77	0.34–1.76		1.35	0.62–2.93		1.29	0.47–3.52		0.84	0.31–2.27		0.88	0.30–2.56		1.44	0.47–4.37	
40–49 years	0.91	0.40–2.09		0.89	0.40–2.00		0.77	0.25–2.44		1.09	0.43–2.78		1.01	0.34–2.97		1.55	0.52–4.56	
50–59 years	1.24	0.58–2.67		1.77	0.86–3.65		0.97	0.33–2.82		1.80	0.76–4.27		1.13	0.41–3.15		2.68	0.97–7.36	
60–69 years	0.84	0.38–1.87		1.16	0.55–2.47		2.02	0.79–5.19		1.72	0.73–4.10		1.34	0.51–3.53		2.17	0.77–6.08	
≥70 years	0.65	0.27–1.56		2.02	0.97–4.20		1.21	0.43–3.38		2.07	0.87–4.91		1.19	0.43–3.29		3.20	1.16–8.86	
State of occupation			0.20			0.11			0.17			0.09			0.11			0.14
With occupation	1.00			1.00			1.00			1.00			1.00			1.00		
Housewife				1.35	0.90–2.02					1.59	0.99–2.57					1.65	1.01–2.72	
Without occupation	1.49	0.81–2.75		1.62	0.88–2.97		1.61	0.82–3.16		1.96	0.98–3.92		1.78	0.89–3.56		1.62	0.76–3.41	
Years of schooling completed			0.31			<0.01			0.71			0.16			0.15			0.29
Junior or senior high school	1.00			1.00			1.00			1.00			1.00			1.00		
College or university	0.78	0.48–1.26		0.49	0.31–0.77		0.90	0.51–1.58		0.69	0.42–1.15		0.65	0.36–1.17		0.76	0.45–1.27	
Seasons of the survey			0.59			0.38			0.36			0.53			0.75			0.15
Winter (February)	1.00			1.00			1.00			1.00			1.00			1.00		
Summer (August)	0.89	0.58–1.36		1.16	0.82–1.65		1.27	0.77–2.10		1.14	0.76–1.69		1.09	0.66–1.79		0.74	0.49–1.12	
Regions			0.48			0.26			0.70			0.85			0.85			0.58
Hokkaido and Tohoku	1.00			1.00			1.00			1.00			1.00			1.00		
Kanto, Keihin, and Koshinetsu	1.11	0.54–2.29		0.65	0.37–1.12		1.24	0.54–2.80		0.72	0.38–1.36		1.42	0.60–3.34		0.92	0.45–1.89	
Hokuriku and Tokai	1.42	0.65–3.13		0.84	0.44–1.60		1.26	0.50–3.17		0.91	0.44–1.89		1.23	0.46–3.30		1.04	0.45–2.37	
Kinki and Hanshin	1.54	0.67–3.50		0.98	0.54–1.78		0.79	0.28–2.26		0.70	0.34–1.44		1.47	0.54–3.99		1.03	0.47–2.25	
Chugoku and Shikoku	0.72	0.25–2.03		0.71	0.35–1.43		0.52	0.14–2.03		0.66	0.29–1.48		0.81	0.23–2.84		0.93	0.39–2.24	
Kyushu	1.67	0.74–3.75		0.48	0.23–1.03		1.25	0.47–3.32		0.84	0.39–1.84		1.60	0.60–4.29		1.67	0.76–3.68	
Size of the city of residence			0.18			0.35			0.68			0.48			0.33			0.63
18 large cities	0.63	0.35–1.13		1.02	0.68–1.54		1.18	0.64–2.15		0.95	0.59–1.54		1.07	0.58–2.00		1.10	0.67–1.78	
City	1.00			1.00			1.00			1.00			1.00			1.00		
Municipality	1.28	0.68–2.44		0.62	0.33–1.20		1.37	0.62–3.01		0.63	0.30–1.33		1.74	0.84–3.62		0.74	0.35–1.55	
Mental health status			<0.01			<0.01			<0.01			<0.01			<0.01			<0.01
Good	1.00			1.00			1.00			1.00			1.00			1.00		
Poor	3.39	2.21–5.21		3.81	2.68–5.41		2.66	1.60–4.42		4.12	2.75–6.15		2.70	1.63–4.48		3.25	2.15–4.91	

Subject with missing data was excluded from the analysis.
p was calculated by the multiple logistic regression analysis.
Abbreviations: DIS, difficulty initiating sleep; DMS, difficulty maintaining sleep; DRS, difficulty resuming sleep; EMA, early morning awakening; AOR, adjusted odds ratio; CI, confidential interval.

Table 5
Factors associated with insomnia.

	Insomnia						Insomnia with daytime dysfunction					
	Male			Female			Male			Female		
	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*	AOR	95% CI	*p*
Age class			0.69			0.01			0.54			0.10
20–29 years	1.00			1.00			1.00			1.00		
30–39 years	1.47	0.71–3.05		1.11	0.55–2.25		1.34	0.34–5.31		0.66	0.17–2.57	
40–49 years	1.36	0.64–2.92		0.93	0.46–1.90		1.32	0.31–5.66		0.99	0.29–3.30	
50–59 years	1.67	0.81–3.44		1.70	0.89–3.26		1.56	0.42–5.77		1.06	0.33–3.39	
60–69 years	1.20	0.58–2.50		1.49	0.77–2.89		0.59	0.15–2.36		1.16	0.37–3.62	
≥70 years	0.96	0.44–2.08		2.37	1.23–4.58		0.43	0.10–1.83		2.65	0.89–7.85	
State of occupation			0.02			0.31			<0.01			0.38
With occupation	1.00			1.00			1.00			1.00		
Housewife				1.32	0.92–1.89					1.56	0.79–3.07	
Without occupation	1.85	1.08–3.15		1.28	0.74–2.22		4.00	1.50–10.70		1.76	0.70–4.40	
Years of schooling completed			0.20			0.01			0.13			0.11
Junior or senior high school	1.00			1.00			1.00			1.00		
College or university	0.77	0.51–1.15		0.61	0.41–0.90		0.52	0.23–1.21		0.56	0.27–1.15	
Seasons of the survey			0.61			0.55			0.74			0.56
Winter (February)	1.00			1.00			1.00			1.00		
Summer (August)	0.91	0.63–1.31		1.10	0.81–1.50		0.89	0.45–1.76		0.85	0.49–1.47	
Regions			0.12			0.71			0.33			0.65
Hokkaido and Tohoku	1.00			1.00			1.00			1.00		
Kanto, Keihin, and Koshinetsu	1.30	0.69–2.45		0.75	0.46–1.23		2.01	0.60–6.74		0.76	0.31–1.87	
Hokuriku and Tokai	1.74	0.88–3.45		0.81	0.45–1.46		3.01	0.85–10.67		0.81	0.28–2.41	
Kinki and Hanshin	1.67	0.82–3.41		0.87	0.50–1.49		2.08	0.51–8.53		0.91	0.34–2.41	
Chugoku and Shikoku	0.65	0.25–1.68		0.63	0.33–1.20		1.47	0.30–7.26		1.50	0.55–4.08	
Kyushu	1.92	0.95–3.90		0.67	0.36–1.25		0.55	0.09–3.28		0.65	0.20–2.15	
Size of the city of residence			0.21			0.70			0.01			0.33
18 large cities	0.81	0.51–1.29		1.09	0.76–1.57		1.02	0.41–2.54		1.36	0.73–2.53	
City	1.00			1.00			1.00			1.00		
Municipality	1.46	0.84–2.52		0.85	0.50–1.44		3.66	1.52–8.81		0.39	0.11–1.33	
Mental health status			<0.01			<0.01			<0.01			<0.01
Good	1.00			1.00			1.00			1.00		
Poor	3.17	2.20–4.58		3.27	2.40–4.45		5.26	2.59–10.68		7.23	3.97–13.16	

Subject with missing data was excluded from the analysis.
p was calculated by the multiple logistic regression analysis.
Abbreviations: AOR, adjusted odds ratio; CI, confidential interval.

winter were compared, some studies reported a deterioration of sleep in summer [38], whereas others reported such an effect in winter [39]. No previous studies have investigated the seasonality of insomnia in Japan. Japan has clear seasonal changes in weather, with hot summers and cold winters, and it was hypothesized that insomnia would increase in the hot and humid summer. However, no significant associations were found between the symptoms of insomnia and seasonality, or between insomnia and seasonality. It is known that, in Japan, the number of patients with seasonal affective disorder increases in the winter [40]. As insomnia is often a symptom of depression, temperature and humidity alone may not be enough to explain the seasonality of insomnia. The present study was the first to investigate the seasonality of insomnia in Japan, and further studies of this topic will be required.

With regard to the regionality of insomnia in Japan, Kageyama et al. conducted a study of 3600 women living in eight cities in Japan to investigate associations between nighttime road traffic volume and insomnia. They concluded that a level–response relationship exists between the nighttime traffic volume on main roads and the risk of insomnia in subjects living in zones 0–20 m from those roads [41]. In a nationwide epidemiological study on seasonal affective disorders, Imai et al. reported that there was a significant difference in the prevalence of these disorders between the southern and northern regions [42]. The present study found no significant associations between the size of the city of residence and insomnia. Thus, the size of a city alone may not have an impact on insomnia. Future investigations will focus more on the environment surrounding the area of residence and the size of the city of residence.

4.4. Interview method

The present survey adopted an interview method to avoid the limitations of a self-administered questionnaire survey. The advantage of an interview-based method is that it has better reliability in that questions are less restricted, more complex questions can be used, and researchers can confirm the responses on the spot. As the questions in the present study were not simple (particularly those regarding DMS and EMA, because the coexistence of DRS was considered), they had to be conveyed accurately so that the participants would respond appropriately. This is why the interview method was adopted. A limitation of the interview survey was that interviewer technique might create variations in the validity and reliability of the interview. To address this shortcoming as much as possible, a structured interview method was employed in the present study.

4.5. Limitations

The present study had some limitations. First, the data on insomnia were based on self-reporting, which could have biased findings. However, several studies have indicated that self-reported data on sleep status show at least a moderate agreement with data from laboratory studies [43,44]. Second, this study was a cross-sectional investigation and, as such, could not demonstrate causal directions. As the main purpose of this study was to clarify the prevalence of insomnia and associated factors among the general population in Japan, and not to discuss a causal relationship between

them, the goal was achieved. Third, as the present study used a questionnaire and the length of the questionnaire was limited, the questions for diagnosing insomnia could not include entire clinical diagnostic criteria for insomnia from the ICSD-III [45] and DSM-V [46]. As a result, diagnosis of insomnia in the present study may have been different from that based on strict clinical diagnostic criteria. In addition, many other factors, such as the psychiatric diagnosis (lifetime and current), physical comorbidities, socioeconomic status such as monthly income, other lifestyle factors, and use of sedatives were not assessed. In future studies, the above-mentioned factors must be considered. Fourth, the response rate of 54.2% in the present study may suggest an existence of non-response bias.

5. Conclusion

This was an epidemiological study of the general nationwide population in Japan regarding insomnia. It is believed that the results convey a more clinically accurate picture of insomnia conditions than data obtained in previous epidemiological studies because of the fact that the definition of insomnia was closer to the clinical diagnostic criteria (eg, it considered DRS and daytime dysfunction).

Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2016.05.013.

References

[1] Fernandez-Mendoza J, Vgontzas AN. Insomnia and its impact on physical and mental health. Curr Psychiatry Rep 2013;15:418.
[2] Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002;6:97–111.
[3] Doi Y, Minowa M, Okawa M, et al. Prevalence of sleep disturbance and hypnotic medication use in relation to sociodemographic factors in the general Japanese adult population. J Epidemiol 2000;10:79–86.
[4] Kim K, Uchiyama M, Okawa M, et al. An epidemiological study of insomnia among the Japanese general population. Sleep 2000;23:41–7.
[5] Ohayon MM, Krystal A, Roehrs TA, et al. Using difficulty resuming sleep to define nocturnal awakenings. Sleep Med 2010;11:236–41.
[6] Wehr TA, Moul DE, Barbato G, et al. Conservation of photoperiod-responsive mechanisms in humans. Am J Physiol 1993;265:R846–57.
[7] APA. Diagnostic and statistical manual of mental disorders (DSM-IV). 4th ed. Washington, DC: American Psychiatric Association; 1994.
[8] Ohayon MM, Caulet M, Priest RG, et al. DSM-IV and ICSD-90 insomnia symptoms and sleep dissatisfaction. Br J Psychiatry 1997;171:382–8.
[9] Goldberg DP, Rickels K, Downing R, et al. A comparison of two psychiatric screening tests. Br J Psychiatry 1976;129:61–7.
[10] Doi Y, Minowa M. Factor structure of the 12-item General Health Questionnaire in the Japanese general adult population. Psychiatry Clin Neurosci 2003;57:379–83.
[11] Suzuki H, Kaneita Y, Osaki Y, et al. Clarification of the factor structure of the 12-item General Health Questionnaire among Japanese adolescents and associated sleep status. Psychiatry Res 2011;188:138–46.
[12] Shimbo M, Nakamura K, Jing Shi H, et al. Green tea consumption in everyday life and mental health. Public Health Nutr 2005;8:1300–6.
[13] Kaneita Y, Ohida T, Osaki Y, et al. Insomnia among Japanese adolescents: a nationwide representative survey. Sleep 2006;29:1543–50.
[14] Kaneita Y, Ohida T, Uchiyama M, et al. The relationship between depression and sleep disturbances: a Japanese nationwide general population survey. J Clin Psychiatry 2006;67:196–203.
[15] Munezawa T, Kaneita Y, Osaki Y, et al. The association between use of mobile phones after lights out and sleep disturbances among Japanese adolescents: a nationwide cross-sectional survey. Sleep 2011;34:1013–20.
[16] Munezawa T, Kaneita Y, Osaki Y, et al. Nightmare and sleep paralysis among Japanese adolescents: a nationwide representative survey. Sleep Med 2011;12:56–64.
[17] Itani O, Kaneita Y, Ikeda M, et al. Disorders of arousal and sleep-related bruxism among Japanese adolescents: a nationwide representative survey. Sleep Med 2013;14:532–41.
[18] Morioka H, Itani O, Kaneita Y, et al. Associations between sleep disturbance and alcohol drinking: a large-scale epidemiological study of adolescents in Japan. Alcohol 2013;47:619–28.
[19] Morioka H, Itani O, Kaneita Y, et al. Factors affecting unhappiness at school among Japanese adolescents: an epidemiological study. PLoS ONE 2014;9:e111844.
[20] Aritake S, Kaneita Y, Ohtsu T, et al. Prevalence of fatigue symptoms and correlations in the general adult population. Sleep Biol Rhythms 2015;13:146–54.
[21] APA. Diagnostic and statistical manual of mental disorders (DSM-III). 3rd ed. Washington, DC: American Psychiatric Association; 1987.
[22] American Academy of Sleep Medicine. International classification of sleep disorders: diagnostic and coding manual (ICSD). Rochester (MN): American Sleep Disorders Association; 1990.
[23] Lichstein KL, Durrence HH, Riedel BW, et al. Epidemiology of sleep: age, gender, and ethnicity. Mahwah (NJ): Lawrence Erlbaum; 2004.
[24] Ohayon MM, Carskadon MA, Guilleminault C, et al. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004;27:1255–73.
[25] Ohayon MM. Prevalence of DSM-IV diagnostic criteria of insomnia: distinguishing insomnia related to mental disorders from sleep disorders. J Psychiatr Res 1997;31:333–46.
[26] Ohayon MM, Zulley J, Guilleminault C, et al. How age and daytime activities are related to insomnia in the general population: consequences for older people. J Am Geriatr Soc 2001;49:360–6.
[27] Vitiello MV, Moe KE, Prinz PN. Sleep complaints cosegregate with illness in older adults: clinical research informed by and informing epidemiological studies of sleep. J Psychosom Res 2002;53:555–9.
[28] Sekine M, Chandola T, Martikainen P, et al. Explaining social inequalities in health by sleep: the Japanese civil servants study. J Public Health (Oxf) 2006;28:63–70.
[29] Friedman EM, Love GD, Rosenkranz MA, et al. Socioeconomic status predicts objective and subjective sleep quality in aging women. Psychosom Med 2007;69:682–91.
[30] Mezick EJ, Matthews KA, Hall M, et al. Influence of race and socioeconomic status on sleep: Pittsburgh Sleep SCORE project. Psychosom Med 2008;70:410–16.
[31] Baglioni C, Battagliese G, Feige B, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011;135:10–19.
[32] Perlis ML, Giles DE, Buysse DJ, et al. Self-reported sleep disturbance as a prodromal symptom in recurrent depression. J Affect Disord 1997;42:209–12.
[33] Buysse DJ, Frank E, Lowe KK, et al. Electroencephalographic sleep correlates of episode and vulnerability to recurrence in depression. Biol Psychiatry 1997;41:406–18.
[34] Dew MA, Reynolds CF 3rd, Houck PR, et al. Temporal profiles of the course of depression during treatment. Predictors of pathways toward recovery in the elderly. Arch Gen Psychiatry 1997;54:1016–24.
[35] Zhang B, Wing YK. Sex differences in insomnia: a meta-analysis. Sleep 2006;29:85–93.
[36] Seeman MV. Psychopathology in women and men: focus on female hormones. Am J Psychiatry 1997;154:1641–7.
[37] Husby R, Lingjaerde O. Prevalence of reported sleeplessness in northern Norway in relation to sex, age and season. Acta Psychiatr Scand 1990;81:542–7.
[38] Ohayon MM, Partinen M. Insomnia and global sleep dissatisfaction in Finland. J Sleep Res 2002;11:339–46.
[39] Pallesen S, Nordhus IH, Nielsen GH, et al. Prevalence of insomnia in the adult Norwegian population. Sleep 2001;24:771–9.
[40] Okawa M, Shirakawa S, Uchiyama M, et al. Seasonal variation of mood and behaviour in a healthy middle-aged population in Japan. Acta Psychiatr Scand 1996;94:211–16.
[41] Kageyama T, Kabuto M, Nitta H, et al. A population study on risk factors for insomnia among adult Japanese women: a possible effect of road traffic volume. Sleep 1997;20:963–71.
[42] Imai M, Kayukawa Y, Ohta T, et al. Cross-regional survey of seasonal affective disorders in adults and high-school students in Japan. J Affect Disord 2003;77:127–33.
[43] Frankel BL, Coursey RD, Buchbinder R, et al. Recorded and reported sleep in chronic primary insomnia. Arch Gen Psychiatry 1976;33:615–23.
[44] Hoch CC, Reynolds CF 3rd, Kupfer DJ, et al. Empirical note: self-report versus recorded sleep in healthy seniors. Psychophysiology 1987;24:293–9.
[45] American Academy of Sleep Medicine. International classification of sleep disorders: diagnostic and coding manual (ICSD-3). 3rd ed. Rochester (MN): American Sleep Disorders Association; 2014.
[46] APA. Diagnostic and statistical manual of mental disorders (DSM-V). 5th ed. Washington, DC: American Psychiatric Association; 2013.

Journal of Epidemiology 27 (2017) 107–111

Contents lists available at ScienceDirect

Journal of Epidemiology

journal homepage: http://www.journals.elsevier.com/journal-of-epidemiology/

Original Article

The association between alcohol use and problematic internet use: A large-scale nationwide cross-sectional study of adolescents in Japan

Hisayoshi Morioka [a], Osamu Itani [b,*], Yoneatsu Osaki [c], Susumu Higuchi [d], Maki Jike [a], Yoshitaka Kaneita [b], Hideyuki Kanda [e], Sachi Nakagome [a], Takashi Ohida [a]

[a] Division of Public Health, Department of Social Medicine, Nihon University School of Medicine, Tokyo, Japan
[b] Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, Yufu, Oita, Japan
[c] Division of Environmental and Preventive Medicine, Department of Social Medicine, Faculty of Medicine, Tottori University, Yonago, Tottori, Japan
[d] National Hospital Organization Kurihama Medical and Addiction Center, Yokosuka, Kanagawa, Japan
[e] Department of Public Health, Faculty of Medicine, Shimane University, Izumo, Shimane, Japan

ARTICLE INFO

Article history:
Received 31 August 2015
Accepted 17 April 2016
Available online 17 January 2017

Keywords:
Internet addiction
Alcohol drinking
Adolescent
Cross-sectional
Japan

ABSTRACT

Background: This study aimed to clarify the associations between the frequency and amount of alcohol consumption and problematic Internet use, such as Internet addiction and excessive Internet use.
Methods: A self-administered questionnaire survey was administered to students enrolled in randomly selected junior and senior high schools throughout Japan, and responses from 100,050 students (51,587 males and 48,463 females) were obtained. Multiple logistic regression analyses were performed in order to examine the associations between alcohol use and problematic Internet, use such as Internet addiction (Young Diagnostic Questionnaire for Internet Addiction ≥5) and excessive Internet use (≥5 h/day).
Results: The results of multiple logistic regression analyses indicated that the adjusted odds ratios for Internet addiction (YDQ ≥5) and excessive Internet use (≥5 h/day) became higher as the number of days in which alcohol had been consumed during the previous 30 days increased. In addition, the adjusted odds ratio for excessive Internet use (≥5 h/day) indicated a dose-dependent association with the amount of alcohol consumed per session.
Conclusions: This study revealed that adolescents showing problematic Internet use consumed alcohol more frequently and consumed a greater amount of alcohol than those without problematic Internet use. These findings suggest a close association between drinking and problematic Internet use among Japanese adolescents.

1. Introduction

In recent years, use of the Internet has increased worldwide, and a large proportion of users are adolescents. In Japan, the internet utilization rate of Japanese adolescents increased from 72.8% in 2001[1] to 96.3% in 2009.[2] With the rapid global popularization of the Internet in recent years, issues related to problematic Internet use (PIU) among adolescents have arisen,[3–11] potentially affecting adolescents' health[3–9,11] and social activities,[3,8–10] including school performance. Recent studies have found that 6.7% of adolescents (aged 15–19 years) in Hong Kong,[5] 4.04% of students (aged 14–18 years) in the United States,[8] and 3.1% of students (mean age, 16.16 years) in Greece[10] were evaluated as Internet addicts. In addition, 17.1% of adolescents (ages 12–18 years) in Singapore reported using the Internet more than 5 h per day.[9]

However, to date, a standard definition of PIU has not been established. Spada et al. argued that there were no officially or broadly accepted diagnostic criteria for PIU, even in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders.[12] Rial et al. reported that previous studies have referred to PIU as Internet addiction (IA), compulsive Internet use, pathological Internet use, excessive Internet use (EIU), or Internet dependence.[13]

Recent studies have reported that PIU is associated with depression and anxiety,[3,5–9,11] sleep disorders (e.g., excessive daytime sleepiness),[4] and substance use (e.g., tobacco).[7,8] PIU among

* Corresponding author. Department of Public Health and Epidemiology, Faculty of Medicine, Oita University, 1-1, Idaigaoka, Hasama-machi, Yufu-shi, Oita 879-5593, Japan.
E-mail address: itaniosamu@oita-u.ac.jp (O. Itani).
Peer review under responsibility of the Japan Epidemiological Association.

http://dx.doi.org/10.1016/j.je.2016.10.004

adolescents is a public health problem that has recently emerged and requires corrective action.

Another public health issue is alcohol use among adolescents. Previous studies have indicated that heavy alcohol consumption during adolescence has long-lasting adverse social and physical effects (e.g., antisocial behavior, such as violence[14]; mental disorders[15]; and a decline in cognitive function[16,17]). Some epidemiological studies have also examined and reported the effects of alcohol use on Internet addiction among adolescents in recent years.[4,18,19] Ko et al. conducted a cross-sectional study of 2114 high school students using a self-administered questionnaire and reported that problematic alcohol use was associated with Internet addiction.[18] Sung et al. performed a nationwide, web-based epidemiological study of adolescents and concluded that alcohol use was significantly associated with Internet addiction only among girls.[19] They postulated two hypotheses to explain the association between alcohol use and Internet addiction. One hypothesis attempted to explain this association from a psychological perspective. Based on Jessor's Problem-Behavior Theory[20] (psychosocial proneness), they viewed problematic forms of behavior, such as alcohol drinking, as a symptom of specific psychosocial factors that are common to many behavioral problems. They inferred that alcohol drinking caused other types of problematic behavior, such as Internet addiction, through such common psychosocial factors. The other hypothesis was based on a neurophysiological perspective: neurotransmitter systems (e.g., serotonergic and dopaminergic) have pathophysiological associations with both alcohol use and behavioral addictions.[21]

The results of these previous epidemiological studies have not contradicted these hypotheses. However, these studies had some important limitations: they were poorly representative national samples, with comparatively small sample sizes of a few thousand subjects[4,18]; they had unknown or extremely low (less than 50%) response rates[4,18]; there was a possibility of substantially large selection bias due to the adoption of a web-based survey method[19]; and they failed to adjust for some potential confounding factors (e.g., lifestyle habits, such as smoking and exercise, as well as mental health status).[4,18] In addition, as none of these studies considered the amount and frequency of alcohol consumption, the existence of a dose-response relationship between alcohol use and Internet addiction was unknown. Therefore, we designed the present study to address these shortcomings and yield more epidemiologically accurate evidence. If the association between drinking and PIU status in adolescents is investigated using a design that compensates for the limitations of the previous studies, the key factors of this association may be elucidated. This knowledge is highly beneficial for public health, not only because it will clarify how drinking and PIU are associated, but also because it will improve healthcare guidance for adolescents who suffer from PIU by addressing their drinking status. Therefore, in the present study, we conducted a survey on alcohol use and PIU (e.g., IA) among junior and senior high school students in Japan.

2. Methods

In the Japanese education system, a child enters junior high school at the age of 12 years. Junior high school is compulsory 3-year education, and those who wish to further their education may go to senior high school for another 3 years. In this report, the grades in junior high school are described as the 7th to 9th grades, and those in senior high school as the 10th to 12th grades.

The number of all registered high schools as of May 2011 in Japan was 14,621 (10,018 junior and 4603 senior high schools). Among them, 140 junior (enrollment: 65,053) and 124 senior high schools (enrollment: 101,591) were randomly selected using a single-stage cluster sampling method with probability proportional to size. We requested the selected schools to participate on a whole-school basis.

The same number of self-administered anonymous questionnaires as the number of students enrolled in a school were sent to all selected schools. Upon arrival of the questionnaires, the responsible person at a given school determined whether the school would participate in the survey on a whole-school basis. At participating schools, the questionnaires were distributed to students by homeroom teachers. Each completed questionnaire was placed in an envelope, and then it was sealed with an adhesive flap by each student. All questionnaires were collected by teachers before being returned to us on a whole-school basis. The survey was conducted between October 2012 and March 2013. The following ethical considerations were taken account: we requested the teachers not to peer at questionnaires over students' shoulders while students were filling out the questionnaires, and we assured anonymity of the questionnaires. This study was approved by the Ethics Committee of the Nihon University School of Medicine.

The questionnaire items used in this study were developed from questionnaires used in similar previous studies. To assess Internet use, all questions of the 8-item version of the Young Diagnostic Questionnaire for Internet Addiction (YDQ)[22] were translated into Japanese and added to the current questionnaire. Reliability and validity of the YDQ were verified in previous studies.[23,24] According to a meta-analysis by Frangos et al., the overall Cronbach's alpha of the YDQ is 0.889.[23] Cronbach's alpha of the YDQ in this study was 0.937. Similar to previous studies involving adolescents,[5,10,25] IA was defined as an affirmative answer to at least 5 of the 8 YDQ questions. We also included a question on the average number of hours spent using the Internet per weekday during the previous 30 days. EIU was defined as 5 or more hours of use per day, based on a previous study with adolescents. As IA and EIU were reported as PIU in a previous study,[13] we referred to both IA and EIU as PIU in the present study. With regard to alcohol consumption, we included questions on the number of days during the past 30 days on which the respondent had consumed alcohol and on the amount of alcohol consumed per session.

As a mental health indicator, the Japanese version of the 12-item General Health Questionnaire (GHQ-12)[26,27] was used. Doi et al. investigated its reliability and validity and reported that the Cronbach's alpha coefficients were 0.83 for men and 0.85 for women. We used the questions from the “depression and anxiety” and “decrease in positive feeling” factors of the GHQ-12.[27] One question from each of the two factors was selected for the present study, and subjects who answered affirmatively to either question were defined as having poor mental health (GHQ score $\geq$1). Previous studies have shown that evaluation of mental health status using depression symptoms with the GHQ-12 and with this cutoff point had a sensitivity of 87.0% and a specificity of 85.1%.[28] The Cronbach's alpha in this study was 0.816.

First, the prevalence of IA based on sex and grade was calculated. Then, the association between sex and IA was examined using a chi-square test. Calculations and analyses of EIU were performed using the same approach.

Next, the prevalence of IA based on the number of days on which alcohol was consumed was calculated. Then, an association between the number of days of alcohol use and IA was examined using a Mantel-Haenszel test for trend. Furthermore, the prevalence of IA based on the amount of alcohol consumed per drinking session was calculated. Then, an association between the amount of alcohol and IA was examined using a Mantel-Haenszel test for trend. Calculations and analyses of EIU were performed using the same approach.

Next, we performed multiple logistic regression analyses to calculate the adjusted odds ratios (aORs) for adolescents who were addicted to the Internet or who had EIU (≥5 h/day) for both the number of days alcohol was consumed and the amount of alcohol consumed per session during the previous 30 days. Regarding the covariates, besides basic attributes (sex and age), eating habits (eating breakfast), and school life (intention to study at university and participating in extracurricular activities), the following items, which are known to have significant associations with IA or EIU, were used: sleep[4] (bedtimes), smoking,[7,8] and mental health.[3,5–9,11] All analyses were performed using SPSS 20.0 for Windows (IBM Corp, Armonk, NY, USA).

3. Results

Of the 140 junior and 124 senior high schools that were selected, 94 (67.1%) and 85 (68.5%) participated, respectively. The total number of enrolled students in the participating junior and senior high schools was 109,847. Of those, 38,871 junior and 62,263 senior high school students responded. The eventual response rate was 60.7% (101,134 responses out of 166,644 eligible students across all initially selected junior and senior high schools). From the collected questionnaires, 1084 were excluded because the respondent's sex or grade was not specified or the responses were inconsistent. Data from the remaining 100,050 questionnaires (51,587 males and 48,463 females) were analyzed.

IA and EIU prevalence with regard to sex is shown in Table 1. The prevalence of IA was 8.1% (95% confidence interval [CI], 7.9%–8.3%) in all participants, 6.4% (95% CI, 6.2%–6.6%) in boys, and 9.9% (95% CI, 9.6%–10.2%) in girls. The result of a chi-square test indicated a significant sex-based difference in the prevalence of IA ($P < 0.001$). The prevalence of EIU was 12.6% (95% CI, 12.4%–12.8%) in all participants, 12.3% (95% CI, 12.0%–12.6%) in boys, and 13.0% (95% CI, 12.7%–13.3%) in girls. The result of a chi-square test indicated a significant sex-based difference in the prevalence of EIU ($P < 0.001$).

IA and EIU prevalence with regard to the number of days on which alcohol had been consumed and the amount of alcohol consumed per drinking session is shown in Table 2. Significant associations were observed between the prevalence of IA and EIU and the number of days on which alcohol had been consumed ($P < 0.001$): with an increase in the number of days of alcohol use, the prevalence of IA and EIU also increased. Similar associations were observed between the prevalence of IA and EIU and the amount of alcohol consumed per drinking session ($P < 0.001$).

Table 3 shows the results of multiple logistic regression analyses using IA and EIU as objective variables, and the number of days on which alcohol was consumed and mental health status as covariates. Significant associations were observed between both the number of days of alcohol use and mental health status and IA. In particular, the aOR odds ratio for IA significantly increased as the number of days of alcohol use increased. Significant associations were also observed between both the number of days on which alcohol was consumed and mental health status and EIU. Similar to IA, the aOR for EIU significantly increased as the number of days of alcohol use increased.

Table 4 shows the results of multiple logistic regression analyses using IA and EIU as objective variables and the amount of alcohol consumed per drinking session and mental health status as covariates. Significant associations were observed between both the amount of alcohol and mental health status and IA. In particular, the aOR for IA significantly increased as the amount of alcohol increased. Significant associations were also observed between both the amount of alcohol consumed per drinking session and mental health status and EIU. Similar to IA, the aOR for EIU significantly increased as the amount of alcohol increased.

4. Discussion

The present study provides evidence of a strong association between alcohol use (amount and frequency) and PIU, with a dose-response relationship. We have inferred some reasons for the observed association of alcohol use with Internet use among adolescents. One possibility is that alcohol consumption and PIU may be different symptoms of a single underlying disorder or disease. PIU is considered a type of behavioral addiction.[29,30] Gambling addiction, another type of behavioral addition, has been found to be associated with substance use.[31,32] An experimental study[33] has shown that PIU and substance dependency involve activation of the same sites in the brain. Moreover, it has been reported that PIU and substance dependency are phenomenologically linked.[34] In addition, a longitudinal study reported on cases of female adolescents who simultaneously developed compulsive Internet use and substance use.[35] The present results also suggest that alcohol use and PIU are symptoms of a single disorder or disease. Adolescents who had impaired emotional control may have dealt with this disability by going on the Internet or using substances. Meanwhile, there is an argument about whether Internet addiction is an addiction to the Internet or to content mediated by the Internet.[36] Moreover, a consensus has not been reached on the definition of PIU itself. Further studies on PIU are required.

Another possibility is the existence of certain disorders or diseases that may be associated with both drinking and PIU, such as insomnia. The prevalence of insomnia symptoms (i.e., difficulty initiating sleep, difficulty maintaining sleep, and early morning awakening) has been reported to be higher among adolescents who are addicted to the Internet than among adolescents who are not addicted.[4] Moreover, insomnia in adolescents has been associated with alcohol use.[37,38] In addition to insomnia, aggressive behavior

Table 1
Prevalence of Internet addiction and exessive Internet use among Japanese adolescents, based on sex.

	Internet addiction[a]			Excessive Internet use (≥5 h/day)[b]		
	Prevalence (%)	95% CI (%)	*P*	Prevalence (%)	95% CI (%)	*P*
Boys	6.4	6.2–6.6	<0.001[c]	12.3	12.0–12.6	<0.001[d]
Girls	9.9	9.6–10.2		13.0	12.7–13.3	
Total	8.1	7.9–8.3		12.6	12.4–12.8	

CI, confidence interval.
Subject with missing data were excluded from the analysis.
[a] Internet addiction: affirmative response to at least 5 or 8 questions from Young Diagnostic Questionnaire for Internet addiction.
[b] Excessive Internet use (≥5 h/day): average of ≥5 h of Internet use per weekday during the previous 30 days.
[c] P was calculated by $\chi 2$ test, Internet addiction × Sex.
[d] P was calculated by $\chi 2$ test, Excessive Internet use × Sex.

Table 2
Prevalence of Internet addiction and exessive Internet use among Japanese adolescents, based on the number of days on which alcohol had been consumed and amount of alcohol consumed per drinking session.

	Internet addiction[a]			Excessive Internet use (≥5 h/day)[b]		
	Prevalence (%)	95% CI (%)	*P for trend*	Prevalence (%)	95% CI (%)	*P for trend*
Number of days on which alcohol had been consumed			<0.001[c]			<0.001[e]
None	7.3	7.1–7.5		10.8	10.6–11.0	
1–5 days	13.2	12.5–13.9		24.4	23.6–25.2	
6–19 days	15.6	13.7–17.5		34.8	32.3–37.3	
≥20 days	25.1	20.5–29.7		41.9	36.6–47.2	
Amount of alcohol consumed per drinking session			<0.001[d]			<0.001[f]
Not drinking	6.5	6.3–6.7		9.4	9.2–9.6	
Less than a glass	9.4	9.0–9.8		11.5	11.0–12.0	
1 or 2 glasses	11.8	11.2–12.4		21.1	20.3–21.9	
3–5 glasses	12.8	11.8–13.8		30.9	29.5–32.3	
≥6 glasses	16.3	14.7–17.9		38.4	36.2–40.6	

CI, confidence interval.
Subject with missing data were excluded from the analysis.
[a] Internet addiction: affirmative response to at least 5 or 8 questions from Young Diagnostic Questionnaire for Internet addiction.
[b] Excessive Internet use (≥5 h/day): average of ≥5 h of Internet use per weekday during the previous 30 days.
[c] P was calculated by Mantel-Haenszel test for trend, Internet addiction × Number of days on which alcohol had been consumed.
[d] P was calculated by Mantel-Haenszel test for trend, Internet addiction × Amount of alcohol consumed per drinking session.
[e] P was calculated by Mantel-Haenszel test for trend, Excessive Internet use (≥5 h) × Number of days on which alcohol had been consumed.
[f] P was calculated by Mantel-Haenszel test for trend, Excessive Internet use (≥5 h) × Amount of alcohol consumed per drinking session.

Table 3
Multiple logistic regression analyses regarding associations between number of days on which alcohol had been consumed and Internet addiction or excessive Internet use.

	Internet addiction[a]			Excessive Internet use (≥5 h/day)[b]		
	AOR	95% CI	*P value*	AOR	95% CI	*P value*
Number of days on which alcohol had been consumed			<0.001			<0.001
None	1.00			1.00		
1-5 days	1.49	1.39–1.60		1.85	1.75–1.95	
6-19 days	1.62	1.38–1.90		2.42	2.13–2.74	
≥20 days	3.36	2.57–4.41		3.55	2.79–4.51	
Mental health status			<0.001			<0.001
Good	1.00			1.00		
Poor	3.25	3.09–3.42		1.36	1.31–1.42	

AOR, adjusted odds ratio; CI, confidence interval.
Adjusted for Sex, Grade, Having breakfast, Bedtimes, Intending to study at university, Participating in extracarricular activities, Smoking.
[a] Internet addiction: affirmative response to at least 5 or 8 questions from Young Diagnostic Questionnaire for Internet addiction.
[b] Excessive Internet use (≥5 h/day): average of ≥5 h of Internet use per weekday during the previous 30 days.

Table 4
Multiple logistic regression analyses regarding associations between amount of alcohol consumed per drinking session and Internet addiction or excessive Internet use.

	Internet addiction[a]			Excessive Internet use (≥5 h/day)[b]		
	AOR	95% CI	*P*	AOR	95% CI	*P*
Amount of alcohol consumed per drinking session			<0.001			<0.001
Not drinking	1.00			1.00		
Less than a glass	1.33	1.25–1.42		1.19	1.12–1.25	
1 or 2 glasses	1.46	1.36–1.57		1.91	1.81–2.02	
3-5 glasses	1.44	1.30–1.60		2.62	2.42–2.83	
≥6 glasses	1.74	1.51 2.01		3.25	2.92 3.63	
Mental health status			<0.001			<0.001
Good	1.00			1.00		
Poor	3.22	3.06–3.39		1.34	1.29–1.40	

AOR, adjusted odds ratio; CI, confidence interval.
Adjusted for Sex, Grade, Having breakfast, Bedtimes, Intending to study at university, Participating in extracarricular activities, Smoking.
[a] Internet addiction: affirmative response to at least 5 or 8 questions from Young Diagnostic Questionnaire for Internet addiction.
[b] Excessive Internet use (≥5 h/day): average of ≥5 h of Internet use per weekday during the previous 30 days.

(e.g., serious fights and carrying weapons) have been associated with PIU[8] and alcohol use[39] among adolescents. Therefore, future research should consider the effects of these confounding factors when examining the association between Internet and alcohol use among adolescents.

The present study had three main strengths. First, the survey used a nationally representative random sample of over 100,000 students. Second, as both the amount and frequency of alcohol consumption were assessed, analyses of the effects of drinking on PIU, including the dose-response relationship, could be performed. Third, in analyses of the associations of alcohol use and Internet addiction, multivariate analyses were used and covariate factors other than alcohol use and Internet addiction were taken into account and appropriately adjusted.

The present study had some limitations. First, because a cross-sectional design was used, a causal relationship could not be determined. For example, this study found associations between alcohol drinking and PIU, such that alcohol drinking may induce PIU. However, the possibility that PIU may induce alcohol drinking cannot be refuted. The latter possibility was suggested in a study by Sun et al. of 1761 high school students in China and 1182 counterparts in the USA.[35] This longitudinal study, with a follow-up survey conducted 1 year later, investigated the association between compulsive Internet use and binge drinking. The authors reported that baseline compulsive Internet use was a significant factor for a change in binge drinking, but they did not find an inverse association. To obtain accurate evidence of causal relationships between alcohol use and Internet addiction, more longitudinal studies with better measures of Internet addiction are needed. Second, a non-response bias may have occurred, as certain schools and students chose not to participate. Moreover, data from long-term absentees are not reflected in the current analyses. Third, the use of a self-administered questionnaire may have resulted in a reporting bias. For instance, data on the number of days alcohol had been consumed and the amount consumed may have been subject to recall bias. Moreover, because underage drinking is illegal in Japan, the respondents may have tended to under-report their alcohol consumption. Fourth, as residential area data of participants were not available, we could not analyze geographical differences.

In conclusion, this study suggests a close association between alcohol drinking and PIU among Japanese adolescents. When addressing the issue of PIU among adolescents, the existence of other forms of problematic behavior, including alcohol drinking, must be considered.

Conflicts of interest

None declared.

Acknowledgments

This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare (H22-junkankitou-shitei-020).

References

1. Communications Usage Trend Survey in 2001 compiled, 2. [cited 2015 Jul 16] Available from: http://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/eng_tsusin_riyou02.pdf.
2. Communications Usage Trend Survey in 2012 Compiled, 6. [cited 2015 Jul 16] Available from: http://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/.
3. Vanr AJ, Kuss DJ, Griffiths MD, Shorter GW, Schoenmakers MT, VDM D. The (co-)occurrence of problematic video gaming, substance use, and psychosocial problems in adolescents. *J Behav Addict*. 2014;3:157–165.
4. Choi K, Son H, Park M, et al. Internet overuse and excessive daytime sleepiness in adolescents. *Psychiatry Clin Neurosci*. 2009;63:455–462.
5. Fu KW, Chan WS, Wong PW, Yip PS. Internet addiction: prevalence, discriminant validity and correlates among adolescents in Hong Kong. *Br J Psychiatry*. 2010;196:486–492.
6. Ko CH, Yen JY, Chen CS, Yeh YC, Yen CF. Predictive values of psychiatric symptoms for internet addiction in adolescents: a 2-year prospective study. *Arch Pediatr Adolesc Med*. 2009;163:937–943.
7. Lee YS, Han DH, Kim SM, Renshaw PF. Substance abuse precedes Internet addiction. *Addict Behav*. 2013;38:2022–2025.
8. Liu TC, Desai RA, Krishnan-Sarin S, Cavallo DA, Potenza MN. Problematic Internet use and health in adolescents: data from a high school survey in Connecticut. *J Clin Psychiatry*. 2011;72:836–845.
9. Mythily S, Qiu S, Winslow M. Prevalence and correlates of excessive Internet use among youth in Singapore. *Ann Acad Med Singap*. 2008;37:9–14.
10. Stavropoulos V, Alexandraki K, Motti-Stefanidi F. Recognizing internet addiction: prevalence and relationship to academic achievement in adolescents enrolled in urban and rural Greek high schools. *J Adolesc*. 2013;36:565–576.
11. Yen JY, Ko CH, Yen CF, Wu HY, Yang MJ. The comorbid psychiatric symptoms of Internet addiction: attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility. *J Adolesc Health*. 2007;41:93–98.
12. Spada MM. An overview of problematic internet use. *Addict Behav*. 2014;39:3–6.
13. Rial Boubeta A, Gomez Salgado P, Isorna Folgar M, Araujo Gallego M, Varela Mallou J. PIUS-a: problematic internet use scale in adolescents. Development and psychometric validation. *Adicciones*. 2015;27:47–63.
14. Swahn MH, Simon TR, Hammig BJ, Guerrero JL. Alcohol-consumption behaviors and risk for physical fighting and injuries among adolescent drinkers. *Addict Behav*. 2004;29:959–963.
15. Marmorstein NR. Longitudinal associations between alcohol problems and depressive symptoms: early adolescence through early adulthood. *Alcohol Clin Exp Res*. 2009;33:49–59.
16. Ehlers CL, Criado JR. Adolescent ethanol exposure: does it produce long-lasting electrophysiological effects? *Alcohol*. 2010;44:27–37.
17. Hanson KL, Medina KL, Padula CB, Tapert SF, Brown SA. Impact of adolescent alcohol and drug use on neuropsychological functioning in young adulthood: 10-year outcomes. *J Child Adolesc Subst Abuse*. 2011;20:135–154.
18. Ko CH, Yen JY, Yen CF, Chen CS, Weng CC, Chen CC. The association between Internet addiction and problematic alcohol use in adolescents: the problem behavior model. *Cyberpsychol Behav*. 2008;11:571–576.
19. Sung J, Lee J, Noh HM, Park YS, Ahn EJ. Associations between the risk of internet addiction and problem behaviors among Korean adolescents. *Korean J Fam Med*. 2013;34:115–122.
20. Jessor R. Problem-behavior theory, psychosocial development, and adolescent problem drinking. *Br J Addict*. 1987;82:331–342.
21. Grant JE, Potenza MN, Weinstein A, Gorelick DA. Introduction to behavioral addictions. *Am J Drug Alcohol Abuse*. 2010;36:233–241.
22. Young KS. Internet addiction: the emergence of a new clinical disorder. *Cyber Psychol Behav*. 1998;1:237–244.
23. Frangos CC, Frangos CC, Sotiropoulos I, eds. *A Meta-analysis of the Reliability of Young's Internet Addiction Test. World Congress on Engineering*. 2012.
24. Laconi S, Rodgers RF, Chabrol H. The measurement of Internet addiction: a critical review of existing scales and their psychometric properties. *Comput Hum Behav*. 2014;41:190–202.
25. Johansson A, Gotestam KG. Internet addiction: characteristics of a questionnaire and prevalence in Norwegian youth (12-18 years). *Scand J Psychol*. 2004;45:223–229.
26. D'Arcy C, Siddique CM. Psychological distress among Canadian adolescents. *Psychol Med*. 1984;14:615–628.
27. Doi Y, Minowa M. Factor structure of the 12-item general health questionnaire in the Japanese general adult population. *Psychiatry Clin Neurosci*. 2003;57:379–383.
28. Suzuki H, Kaneita Y, Osaki Y, et al. Clarification of the factor structure of the 12-item General Health Questionnaire among Japanese adolescents and associated sleep status. *Psychiatry Res*. 2011;188:138–146.
29. Beard KW, Wolf EM. Modification in the proposed diagnostic criteria for Internet addiction. *Cyberpsychol Behav*. 2001;4:377–383.
30. Holden C. 'Behavioral' addictions: do they exist? *Science*. 2001;294:980–982.
31. Griffiths M, Parke J, Wood R. Excessive gambling and substance abuse: is there a relationship? *J Subst Use*. 2002;7:187–190.
32. Griffiths M, Sutherland I. Adolescent gambling and drug use. *J Community & Appl Soc Psychol*. 1998;8:423–427.
33. Ko CH, Liu GC, Hsiao S, et al. Brain activities associated with gaming urge of online gaming addiction. *J Psychiatr Res*. 2009;43:739–747.
34. Liu T, Potenza MN. Problematic Internet use: clinical implications. *CNS Spectr*. 2007;12:453–466.
35. Sun P, Johnson CA, Palmer P, et al. Concurrent and predictive relationships between compulsive internet use and substance use: findings from vocational high school students in China and the USA. *Int J Environ Res Public Health*. 2012;9:660–673.
36. Griffiths M. Does internet and computer "addiction" exist? Some case study evidence. *CyberPsychol Behav*. 2000;3:211–218.
37. Kaneita Y, Ohida T, Osaki Y, et al. Insomnia among Japanese adolescents: a nationwide representative survey. *Sleep*. 2006;29:1543–1550.
38. Morioka H, Itani O, Kaneita Y, et al. Associations between sleep disturbance and alcohol drinking: a large-scale epidemiological study of adolescents in Japan. *Alcohol*. 2013;47:619–628.
39. Chen CY, Storr CL, Tang GM, Huang SL, Hsiao CK, Chen WJ. Early alcohol experiences and adolescent mental health: a population-based study in Taiwan. *Drug Alcohol Depend*. 2008;95:209–218.

Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies

Toshi A Furukawa, Andrea Cipriani*, Lauren Z Atkinson, Stefan Leucht, Yusuke Ogawa, Nozomi Takeshima, Yu Hayasaka, Anna Chaimani, Georgia Salanti*

Summary

Background **Previous studies have shown that placebo response rates in antidepressant trials have been increasing since the 1970s. However, these studies have been based on outdated or limited datasets and have used inappropriate statistical methods. We did a systematic review of placebo-controlled randomised controlled trials of antidepressants to examine associations between placebo-response rates and study and patient characteristics.**

Methods **In this systematic review, we searched for published and unpublished double-blind randomised placebo-controlled trials of first-generation and second-generation antidepressants for acute treatment of major depression in adults (update: Jan 8, 2016). The log-transformed proportions of placebo response, defined as 50% or greater reduction in depression severity score from baseline, were meta-analytically synthesised for each year. We then looked for a structural break point in the secular changes in these characteristics through the years and examined the influence of the study year and other trial and patient characteristics on the response rates through meta-regression.**

Findings **We identified 252 placebo-controlled trials (26 324 patients on placebo) done between 1978 and 2015. There was a structural break in 1991, and since then, the average placebo response rates in antidepressant trials have remained constant in the range between 35% and 40% (relative risk [RR] 1·00, 95% CI 0·97–1·03, p=0·99, for every 5-year increase). The length of the study and the number of study centres were significant factors (RR 1·03, 95% CI 1·01–1·05 for 1 more week in trial length; 1·32, 1·11–1·57 for multicentre *vs* single-centre trials).**

Interpretation **Contrary to the widely held belief, the average placebo response rates in antidepressant trials have been stable for more than 25 years. This new evidence should have an effect on the interpretation of the scientific literature and the future of psychopharmacology, both from a clinical and methodological point of view.**

Funding **Japan Society for Promotion of Science, Great Britain Sasakawa Foundation.**

Lancet Psychiatry 2016

Published Online
October 7, 2016
http://dx.doi.org/10.1016/S2215-0366(16)30307-8

See Online/Comment/
http://dx.doi.org/10.1016/S2215-0366(16)30308-X

*These authors contributed equally

Department of Health Promotion and Human Behavior (Prof T A Furukawa, Y Ogawa MD, N Takeshima MD, Y Hayasaka MD), **Department of Clinical Epidemiology** (Prof T A Furukawa), **Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; Department of Psychiatry, University of Oxford, Oxford, UK** (A Cipriani PhD, L Z Atkinson MSc); **Department of Psychiatry and Psychotherapy, Technische Universität München, Klinikum rechts der Isar, Munich, Germany** (Prof S Leucht MD); **Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland** (A Chaimani PhD, G Salanti PhD); **and Department of Hygiene and Epidemiology, University of Ioannina, Greece** (A Chaimani PhD, G Salanti)

Correspondence to:
Dr Andrea Cipriani, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK
andrea.cipriani@psych.ox.ac.uk

Introduction

It has been widely accepted that placebo response rates in antidepressant trials have been increasing over the past 30 years. The first systematic study[1] on this topic revealed that among the 75 placebo-controlled antidepressant trials up to 2000, there was a positive correlation between the proportion of responders on placebo and the year of publication (r=0·45, 95% CI 0·25–0·61). A very similar association was found in paediatric antidepressant trials (r=0·64, 95% CI 0·02–0·91).[2] Other studies examined pre–post raw change scores[3,4] or their effect sizes[5,6] in the placebo groups and confirmed a significant association between the thus-measured placebo response and the year of study publication.

Since the increasing placebo response was suspected of contributing to increasing numbers of so-called failed antidepressant trials,[7,8] many authors explored factors behind this relation in the hope of finding measures to curtail this increase. Walsh and colleagues[1] found that the length of the trial and the minimum depression severity required at baseline also influenced the placebo response rates in adult antidepressant trials; however, when jointly entered into multiple linear regression, only the year of publication remained significantly associated with the response to placebo.[1] Age, sex, use of placebo run-in or co-medication did not show significant association.[1] The number of recruiting sites and the number of randomised patients were identified as important factors associated with placebo response in paediatric trials.[2] Other patient or study characteristics suggested to be associated with placebo response include baseline severity,[3,6] number of groups or probability of being allocated to placebo,[9,10] dosing schedule,[11] length of trial,[3,12] and inflation of baseline severity.[13] However, all these empirical studies have methodological problems. Many of them are based on samples that are not only out of date now but also biased, because they included only published trials, often only in English language.[1,3,6,12,14] Other reviews included data from the US Food and Drug Administration, which thus precluded publication bias but also limited the number and scope of available studies.[4,5,11] Moreover, in terms of statistical analysis, previous studies examined the association between the proportion of placebo

Panel: Research in context

Evidence before this study

It has been accepted knowledge that placebo response rates in antidepressant trials have been increasing over the past decades since 1970s. This finding has been associated with the increasing number of so-called "failed" antidepressant trials, and many authors have examined and debated factors behind this association, such as depressive symptoms' severity, inflation of baseline score, and the use of placebo run-in phases in the trial design. Unfortunately, previous reviews have been based on incomplete datasets or used inappropriate statistical methods (for instance, the association between the proportion of placebo responders and the year of publication was examined as correlation coefficients or linear regressions without taking into account the precision of estimated proportion in each included study and erroneously assuming normal distribution for proportions). We did a systematic and comprehensive search for published and unpublished randomised trials of first-generation and second-generation antidepressants for acute treatment of adult patients with major depression (update: January, 2016), to examine the association between proportion of placebo-responders and study and patient characteristics.

Added value of this study

We identified a dataset of 252 placebo-controlled double-blind studies done since 1978 (overall, 26 324 participants randomly assigned to placebo). In our large dataset, there was a structural break in 1991, and the average placebo response rate in antidepressant trials has basically remained constant in the range of 35–40% since 1991. We were able to replicate the reported increase in placebo response rates before 2000. However, we found that this analysis was confounded by methodological factors (such as shorter duration of trials and preponderance of single-centre studies), which were frequent in the very early trials and became less often employed since 1990s. Controlling for these variables, the association between placebo response rate and study year was no longer significant.

Implications of all the available evidence

Placebo response rates in antidepressant trials have, on average, stayed constant over the past 25 years. This new evidence should have an effect on the interpretation of the scientific literature and future psychopharmacology, with both clinical and methodological implications. In terms of study design, some important factors should be reconsidered. Appropriately timed assessments and multicentre design are necessary for trials to provide results directly relevant and applicable to real-world clinical practice. Trialists need to use design characteristics to make their trials as clinically relevant as possible.

responders and the year of publication as correlation coefficients or linear regressions without taking into account the precision of estimated proportion in each included study and erroneously assuming normal distribution for proportions.[1,2,14]

We did a comprehensive and systematic search for published and unpublished double-blind placebo-controlled trials of first-generation and second-generation antidepressants (the search was part of an update of a network meta-analysis about antidepressants in major depression).[15] We focused on the placebo response rates and, using the largest dataset so far, we aimed at addressing the following questions: whether the believed increase in placebo response rates had persisted up to 2015, and if not, whether we could replicate the previous findings,[1] and what patient and study characteristics could influence the studied associations.

Methods

Data sources and search criteria

We included all double-blind randomised trials comparing one of the following active drugs with placebo in the acute treatment of major depression: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone, and vortioxetine. We included all the second-generation antidepressants licensed in Europe, USA, Australia, and Japan, and, of the older drugs, we selected the two tricyclics included in the WHO list of essential medicines (amitriptyline and clomipramine), plus trazodone and nefazodone because they had very distinct effect and tolerability profiles.[15] We included randomised trials with patients aged 18 years or older, of both sexes, and with a primary diagnosis of unipolar major depression according to standard operationalised diagnostic criteria. We searched the relevant electronic databases including Cochrane CENTRAL, CINAHL, Embase, LiLACS, MEDLINE, MEDLINE In-Process, and PsycINFO, supplemented by searches for published, unpublished, and ongoing randomised trials in major drug-approving agencies, clinical trial registries, and pharmaceutical company websites up to Jan 8, 2016. We also contacted the National Institute for Health and Care Excellence (NICE, UK), the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG, Germany), and other relevant organisations and experts in the field for any additional information not already identified. For details, please refer to the full study protocol.[15] No date limits or language restrictions were applied to any of the searches.

At least two people (TAF, ACi, LZA, YO, NT, or YH) independently checked reference titles and abstracts thus identified. Full texts of all potentially eligible studies were retrieved, and were further inspected using the same

eligibility criteria. Any disagreements were resolved by discussion or in consultation with a third member of the review team. Two or more people (TAF, ACi, LZA, YO, NT, or YH) then independently extracted data using the pre-defined data extraction sheet. Two or more independent raters also assessed the quality of each study using the Cochrane risk of bias tool.[16] Each study was then classified as having low risk of bias (none of the domains rated at high risk of bias and three or fewer domains rated at unclear risk), moderate risk of bias (one domain rated at high risk of bias or none rated at high risk of bias but four or more at unclear risk), or high risk of bias (all other cases).[15]

As we obtained information from various sources including single trial publications, published meta-analyses, regulatory agencies' documents, and company website information, great efforts were expended to identify and avoid duplication, both before and after data extraction, and to contact original study authors for supplemental information, if needed.

Statistical analyses

We examined the proportions of placebo response in the placebo groups of all the identified trials in the following manner. We first tabulated the proportion of placebo response by half-decades since the earliest trial. The log-transformed proportion of responders was meta-analytically synthesised for each half decade, using the random effects model and the method of moments[17] to estimate the between-study component of variance τ^2 with the *metan* command in Stata version 14.1 (StataCorp, College Station, TX, USA). We defined response as a 50% or greater reduction in the total score from baseline to week 8 on a standardised observer-rating scale for depression. We used the Hamilton Rating Scale for Depression (HRSD) or, if HRSD was not used, another standardised and validated observer-rating scale such as the Montgomery-Asberg Depression Rating Scale (MADRS). Since we were interested in the acute treatment of major depression, assessments had to be done between 4 and 12 weeks, and the ones closest to 8 weeks were prioritised. We excluded longer-term studies from the statistical analyses if they did not provide data for the 4–12 week period. When the number of responders was not reported but baseline mean and endpoint mean and its standard deviation on the depression rating scale were provided, we calculated the number of responders by using a validated imputation method.[18] To abide by the intention-to-treat principle, we used the number of patients randomised as the denominator of our primary analysis and we assumed any participants with an unknown outcome to be non-responders in accordance with a validated method to estimate relative treatment effects.[19]

Next we examined if there was any structural break in this time series—ie, when a time series abruptly changes at a point in time, by applying the command *estat sbsingle* in Stata to our series of meta-analytically pooled response rates for each year from 1978 to 2015. The structural break test helps us to determine when and whether there is a significant change in the data series. We then ran a meta-regression of log of placebo response rate by the year of study completion. We gave preference to this variable over the year of publication because, by definition, the latter is unavailable for unpublished studies. We next ran meta-regression of log of placebo response rates by year of the study completion while controlling for various participant and study design characteristics. We recorded depression severity as measured by different versions of HRSD and MADRS into HRSD-17 scores by using the conversion table based on the item response theory.[20] We

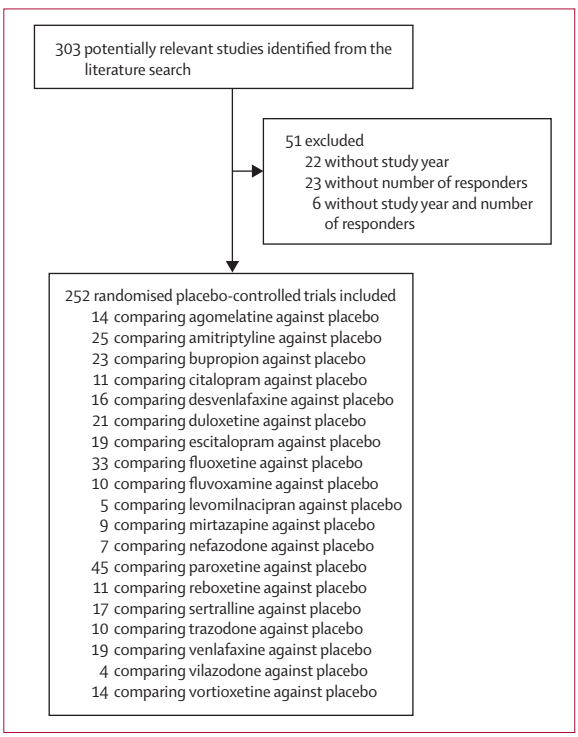


Figure 1: **Study selection**
*The total added number exceeds 252 because 60 trials involved in more than one active drug.

	Number of studies	Range of proportions of responders	Weighted mean proportion of responders (95% CI)	I^2, %	Tau
1978–85	27	0·00–0·41	0·30 (0·27–0·33)	12·0%	0·087
1986–90	32	0·08–0·50	0·28 (0·25–0·31)	32·1%	0·171
1991–95	34	0·14–0·54	0·35 (0·32–0·39)	71·6%	0·232
1996–2000	39	0·21–0·53	0·38 (0·35–0·41)	69·3%	0·181
2001–05	47	0·18–0·58	0·40 (0·37–0·42)	72·8%	0·176
2006–10	46	0·23–0·70	0·37 (0·34–0·40)	78·3%	0·207
2011–15	27	0·17–0·63	0·36 (0·33–0·40)	84·9%	0·232
All years	252	0·00–0·70	0·36 (0·35–0·37)	74·1%	0·211

Table 1: **Proportion of placebo responders by half-decades**

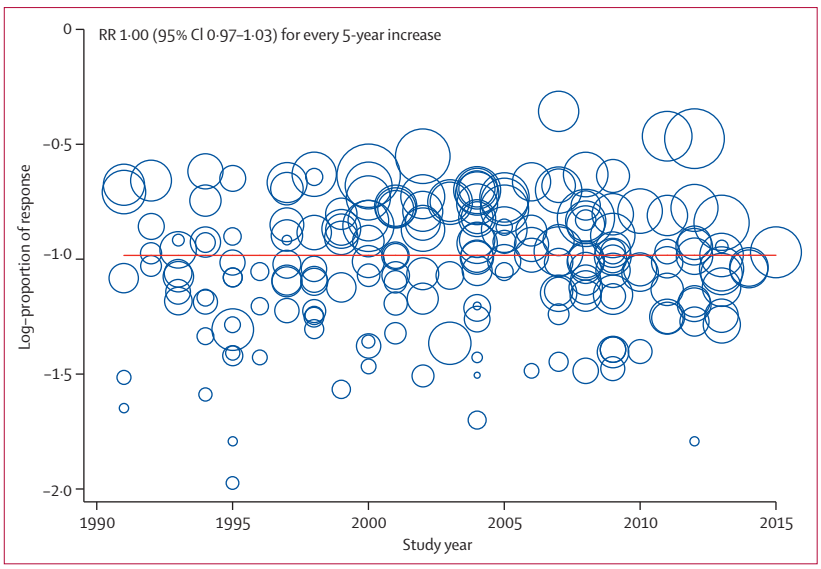


Figure 2: **Meta-regression of log-proportion of response to placebo by study year since 1991 (193 studies)**
RR=relative risk.

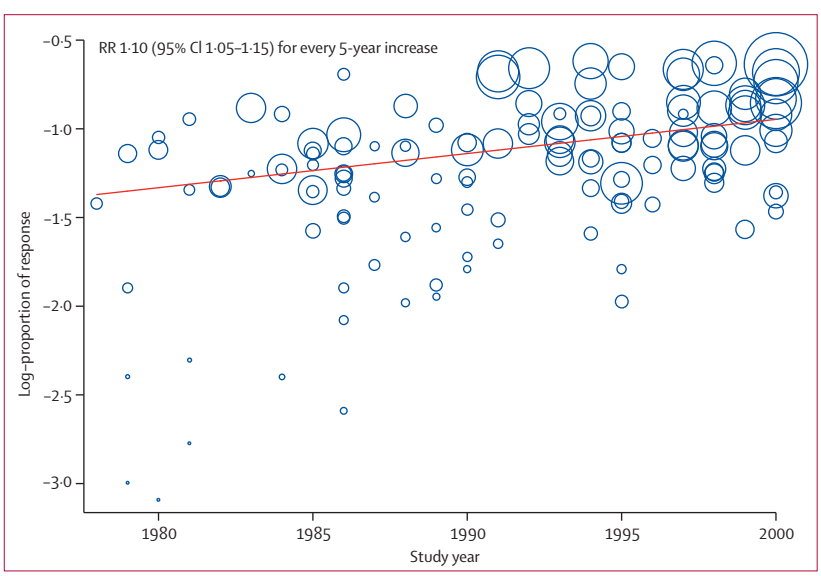


Figure 3: **Meta-regression of log-proportion of response to placebo by study year up to 2000 (132 studies)**
RR=relative risk.

See Online for appendix

entered each variable individually in addition to the study year first, and when the participant or study characteristics proved to exert significant influence, they were entered together to see whether the association between log of placebo response rates and the year of the study was affected. For meta-regression, we used the *metareg* command in Stata with the method of moments estimator for heterogeneity. The percentage of heterogeneity standard deviation that is explained by additional covariates was monitored and reported.

The influential study by Walsh and colleagues[1] was published in 2002 and included trials up to 2000. We examined whether we could replicate their finding by limiting our sample to trials up to the year 2000, while taking a statistically more appropriate dependent variable and employing meta-analytic methods with proper weighting. We first ran simple meta-regression by year of the study and then examined if additional study characteristics affected the association.

We did the following sensitivity analyses to test the robustness of our primary findings: (1) exclusion of studies that imputed the number of responders based on the continuous depression severity scores;[18] (2) use of the number of participants analysed instead of the number of participants randomised as denominator for calculations of the response rate; (3) use of the year of publication instead of the year of study completion; (4) exclusion of unpublished studies; (5) exclusion of studies at high or moderate risk of bias; (6) use of log-odds of placebo response instead of log of placebo response rate as the dependent variable; (7) use of pre–post change on the rating scale as the dependent variable.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. TAF and ACi had full access to all the data in the study and had final responsibility for the decision to submit for publication

Results

The literature search identified 303 placebo-controlled trials. However, in 51 of them either the year of the study or any information about efficacy was missing, so the remaining 252 studies constituted the dataset for our analysis (134 randomised trials with published data only, 74 with both published and unpublished data, and 44 studies with unpublished data only; figure 1). The characteristics of the included studies are summarised in the appendix. Between 1978 and 2015, placebo response rates ranged widely from 0% to 70% but the weighted mean proportion of responders appeared to converge towards 35% to 40%, especially after the early 1990s (table 1). It must be noted that overall the *I*-squared value remained high (74·1% overall), indicating substantial heterogeneity in placebo response rates among the included trials throughout these years.

The structural break test suggested that the break date was the year 1991 (p=0·04), which coincided well with the observed values reported in table 1. When pooled through meta-analysis, the summary response rate for this time period was 0·37 (95% CI 0·36–0·39) with a heterogeneity standard deviation of 0·203. We ran meta-regression to examine the trend after this date and also to explore possible confounders in the association. Figure 2 shows meta-regression of log of proportion of placebo response by the year of the study completion after 1991. Since 1991, the relative risk in the proportion

of placebo response was 1·00 (95% CI 0·97–1·03, p=0·99) for every 5-year increase in the study year. When individual participant or study design characteristics were examined, the length of the study and the number of study centres emerged as significant factors: the longer studies had greater response rates, and the multicentre trials had greater response rates than single-centre studies (RR 1·03, 95% CI 1·01–1·05 for 1 more week in trial length; 1·32, 1·11–1·57 for multicentre trials *vs* single-centre trials). When both variables were entered together in the meta-regression, both remained significant predictors (RR 1·03, 95% CI 1·00–1·05, for trial duration; 1·31, 1·10–1·56 for multicentre trials) whereas the study year remained non-significant (0·99, 0·96–1·01, for every 5-year increase). This multivariable model explained 2·2% of the total heterogeneity standard deviation.

We examined if we could replicate Walsh and colleagues' findings[1] by limiting the studies up to the year 2000. The summary response rate from simple meta-analysis was 0·34 (95% CI 0·32–0·35) with a heterogeneity standard deviation 0·223. Meta-regression of log of proportion of placebo responders by the study year up to 2000 clearly indicated the increasing placebo response (RR 1·10, 95% CI 1·05–1·15, p<0·0001, for every 5-year increase; figure 3, table 2). When possible confounders were examined individually in meta-regression, the number of study centres, the dosing schedule, the length of the trial, and the number of patients randomised emerged as possible confounders (table 2). In addition to the same tendencies observed for the studies after 1991 for the number of weeks and the number of study centres, studies using fixed dose regimens showed lower placebo response rates than those using flexible dose regimens, and the smaller groups tended to show lower response rates. When all these four variables were entered into multivariable meta-regression, the mean response rate for a hypothetical 4-week, single-centre trial with fixed dose regimen in the earliest year of publication (1978) was estimated to be 0·19 (95% CI 0·16–0·23). The response rate increases for a multicentre trial, for a flexible dosing schedule, and for 1 more week in the length in the trial were still significant (table 2). However, the effect of the study year and the number of patients randomised were no longer significant. This multivariable model was able to explain 28·4% of the total heterogeneity and 18·4% of the residual heterogeneity of the meta-regression model with the study year alone. When we examined the secular changes in these characteristics through the years, we found that the studies were much shorter and more often single-centred in the 1980s than in 1990s and later, whereas the flexible dose regimen became increasingly unpopular in the 2000s and later (table 3).

All sensitivity analyses revealed no effect of year among studies after 1991, providing support for robustness of our main findings (appendix).

Discussion

In our study, we found that the average placebo response rate in antidepressant trials has remained constant in the range of 35–40% since 1991, year at which we found a structural break. We were able to replicate the reported increase in placebo response rates before 2000, but this association was no longer significant when controlling for methodological factors (such as shorter duration of trials and preponderance of single-centre studies), which

	Number of studies	Study year RR (for every 5-year increase, 95% CI)	Covariate RR (95% CI) [interpretation]	Tau
Model with year as covariate				
Year	132	1·10 (1·05–1·15)	..	0·196
Models with year and one additional covariate				
Additional covariate				
Age	119	1·10 (1·05–1·15)	1·00 (0·92–1·08) [for every 10-year increase in mean age]	0·198
Sex	57	1·10 (1·02–1·18)	1·03 (1·00–1·07) [for every 10% increase in proportion of women]	0·208
Region	130	1·10 (1·05–1·15)	0·94 (0·77–1·14) [North America *vs* Europe]; 1·03 (0·81–1·31) [cross-continental *vs* Europe]	0·202
Number of study centres	118	1·07 (1·02–1·11)	1·37 (1·21–1·56)[multicentre *vs* single-centre]	0·180
Setting*	66	1·06 (1·00–1·14)	0·86 (0·70–1·06) [secondary or tertiary care *vs* primary care]	0·205
Patient status†	119	1·10 (1·05–1·15)	1·00 (0·87–1·16) [outpatients *vs* inpatients]	0·202
Baseline eligibility threshold‡	108	1·09 (1·04–1·14)	0·96 (0·85–1·07) [for every five point increase in HAMD17 threshold score]	0·193
Baseline severity‡	111	1·09 (1·04–1·14)	0·91 (0·83–1·00) [for every five point increase in baseline HAMD17]	0·181
Number of arms	132	1·10 (1·05–1·15)	1·13 (0·78–1·63) [one arm more]	0·198
Placebo run-in	106	1·10 (1·04–1·15)	0·89 (0·75–1·06) [present *vs* absent]	0·192
Flexible *vs* fixed dose regimen	131	1·10 (1·06–1·15)	1·13 (1·01–1·27) [flexible *vs* fixed]	0·187
Length of trial (in weeks)	124	1·05 (1·00–1·11)	1·05 (1·02–1·08) [1 week more]	0·182
Rescue medication	57	1·07 (0·99–1·15)	0·90 (0·76–1·07) [present *vs* absent]	0·191
Number of randomised patients	132	1·08 (1·03–1·13)	1·11 (1·02–1·22) [100 participants more]	0·196
Multivariable meta-regression (using all significant variables)				
Study year	112§	1·04 (0·99–1·09)§	..	0·160§
Number of study centres	..	..	1·41 (1·22–1·62)	..
Flexible *vs* fixed dose regimen	..	..	1·17 (1·04–1·31)	..
Length of trial	..	..	1·04 (1·01–1·06)	..
Number of randomised patients	..	..	0·97 (0·89–1·07)	..

RR=relative risk. *Studies done in both primary and secondary or tertiary care were classified as ones in primary care. †Studies recruiting both inpatients and outpatients were classified as ones recruiting inpatients. ‡Converted into Hamilton Rating Scale for Depression-17 using the conversion table.[20] §Results for all five variables.

Table 2: **Factors influencing proportion of placebo responders before 2000**

	Number of studies	Mean length of trial, weeks (range)	Multicentre trials, n (%)	Flexible dosing, n (%)
1978–1985	27	4·6 (4–6)	13 (48·2%)	23 (85·2%)
1986–1990	32	5·8 (4–8)	11 (34·4%)	24 (75·0%)
1991–1995	34	6·7 (4–12)	22 (64·7%)	25 (73·5%)
1996–2000	39	8·0 (4–12)	34 (87·2%)	29 (74·4%)
2001–2005	47	7·8 (6–12)	44 (93·6%)	20 (42·6%)
2006–2010	46	7·6 (6–12)	45 (97·8%)	18 (39·1%)
2011–2015	27	8·1 (6–12)	25 (92·6%)	3 (11·1%)
All years	252	7·1 (4–12)	194 (77·0%)	142 (56·4%)

Table 3: **Secular changes in study characteristics of placebo-controlled trials of antidepressants**

were frequent in the very early trials and became less often used since 1990s.

This is the largest systematic review of antidepressant randomised trials done so far and sheds new light in the current heated debate about diminishing efficacy of antidepressant drugs. First, the average placebo response rates have stayed constant since 1991 in the range of 35–40%. Second, the myth of increasing placebo response was due to the exceptionally low rates in 1980s and before, which were confounded especially by the shorter duration of the trials and the preponderance of single-centre studies. These findings are quite contrary to the heretofore almost unanimously held belief in the medical literature that the antidepressant trials are fraught by increasing placebo response rates (interestingly, the meta-analysis by Rief and colleagues[6] found that only doctor-reported outcomes, such as HAMD scores, showed the rise in placebo response of studies published before 2000, but this was not true for patient-reported outcomes).[1–6,14]

Based on our large dataset, however, there is no doubt now that the average placebo-response rates have been staying constant since 1990s, as far as antidepressant trials in depression are concerned. These results are robust since they remained consistent in all the sensitivity analyses. It might well be the case, however, that the association is different in other psychiatric disorders.[21,22]

The second insight may do away with much speculation in the literature about the causes of increasing placebo response. Of the variables found to be significant in the scientific literature, when we examined, in our much larger dataset and with appropriate statistical methods, the length of trial, the number of study centres, the number of randomised patients, the baseline severity, the number of groups, and the dosing schedule, only the length of trial and the number of study centres were found to exert consistent influence independently of each other. Previous findings of the association between increasing placebo response rates and study year became non-significant when we controlled for the confounding factors. In other words, the reported increase before 2000 is a methodological artifact because, when the relevant trial design characteristics were held relatively constant after 1990s, the placebo response rates stayed stable. Publication status did not influence this association since exclusion of unpublished studies did not change the findings (appendix).

There are several strengths of our study. First, we were able to include 252 placebo-controlled studies of antidepressants published between 1978 and 2015, in contrast to the 75 studies included in the study by Walsh and colleagues (up to 2000)[1] or 107 studies included in the study by Undurraga and colleagues (up to 2010).[14] We cannot rule out the possibility that our search still failed to identify some unpublished studies; however, the total number of participants on placebo in the current study was three-to-four times larger than in the previous studies (26 323 *vs* 6285[1] or 9925[14]). Second, we used methods that were more appropriate statistically throughout the analyses by taking the log of response rate instead of raw rates, by weighting the studies by their sample sizes, and by using the objective method to detect the change point in the longitudinal time series.

This study has some limitations. We included trials of an a-priori defined list of first-generation and second-generation antidepressants but we were not able to include all medications known to act as antidepressants to date. However, this approach was inevitable as there is no uniformly accepted list of effective antidepressants and some compounds might be approved for major depression in some countries but not in others. Moreover, by specifying the drugs to focus on, we think we were able to conduct a more systematic and comprehensive search for both published and unpublished trials, which is demonstrated by the size of the current dataset that far exceeds those of any previous study on the topic. To be clinically informative, we included only licensed drugs, therefore we did not include studies of drugs whose development failed. This failure could have happened because of very high response rates. The non-inclusion of placebo response rates from failed development programmes could have biased our sample of eligible studies toward a lower rate of placebo response; however, we carried out a comprehensive search of more than 250 published and unpublished trials and we do not think that the potential omission of a handful of such studies would materially affect the overall results.

The substantial heterogeneity as suggested by high *I*-squared values suggests that the placebo response rates remain highly variable through the years. To address this methodological issue, we therefore used the random effects meta-analytic methods.[23] Only access to individual patient data would allow researchers to better evaluate moderators and mediators of treatment effect; however, this is not always easy and often very time consuming.[24] Moreover, this analysis should be carried out in a systematic way and across multiple drugs, and not limited to a small sample of available studies about one drug, as previously done in the field of antidepressant trials.[25]

When we replicated Walsh and colleagues' findings before 2000, very early studies were on average smaller than those published later. This difference means that in the meta-regression analysis the study year was correlated with study weights, potentially challenging the assumptions of a joint analysis. When the sample size was examined as a univariable explanatory variable, it was a significant predictor but lost its significance when entered together with other covariates in the multivariable model (table 2). We also ran a post-hoc non-weighted multivariable regression, which indicated that neither the study year nor the sample size was a significant predictor, differently from what was previously reported.[26]

In conclusion, even if the average response to placebo has remained constant over the past 25 years, 35–40% is still a high proportion of patients. Assuming this is the best estimate we have of an average response to placebo, important implications for clinicians and researchers can be drawn. The expectation of improvement, classic conditioning, and the contact with a health-care environment with supportive and therapeutic features contribute to the objective response observed in patients with major depression who are randomly assigned to placebo.[27] These non-pharmacological aspects, however, are usually not provided to the same extent in standard clinical practice. Clinicians should create a specific context and level of therapeutic contact, to enhance non-specific effects of treatment and gain greater treatment response.[28] In terms of research, innovation in psychopharmacology is urgently needed not only for drug discovery and development, but also in terms of clinical trials' design. The key question is whether there is still need to have placebo-controlled phase 3 studies.[29] Clinical research organisations have been running trials in depression for most pharmaceutical companies since the early 1990s and they are financially incentivised to have as many visits as possible to increase the duration and therefore the cost per patient of the trial. Moreover, they tend to use as many centres as possible so that the study is completed sooner. To evaluate the next putative antidepressants, future studies should require the development of more efficient study designs to improve signal detection in drug development studies and an increased antidepressant response in clinical treatment.[30] It is time for academics, pharmaceutical industry, and regulators to create new models of drug discovery and drive innovation in the methodology of drug development.[31]

Contributors

TAF and ACi conceived the study and drafted the manuscript. GS, ACi, ACh, and TAF designed the methods. TAF, ACi, LZA, YO, NT, and YH selected the articles and extracted the data. TAF, ACh, and GS analysed the data. TAF, ACi, and SL interpreted the results. All authors critically revised the manuscript. All authors read and met the International Committee of Medical Journal Editors (ICMJE) criteria for authorship and agree with the results and conclusions of this Article.

Declaration of interests

TAF has received lecture fees from Eli Lilly, Janssen, Meiji, MSD, Otsuka, Pfizer, and Tanabe-Mitsubishi, and consultancy fees from Sekisui Chemicals and Takeda Science Foundation; has received royalties from Igaku-Shoin and Nihon Bunka Kagaku-sha publishers; and has received grant or research support from Mochida and Tanabe-Mitsubishi; he is diplomate of the Academy of Cognitive Therapy. ACi was expert witness for Accord Healthcare for a patent issue about quetiapine extended release. SL has received honoraria for lectures from Eli Lilly, Lundbeck (Institute), Pfizer, Janssen, BMS, Johnson and Johnson, Otsuka, Roche, Sanofi-Aventis, ICON, Abbvie, AOP Orphan, and Servier; has received honoraria for consulting or advisory boards from Roche, Janssen, Lundbeck, Eli Lilly, Otsuka, and TEVA; for the preparation of educational material and publications from Lundbeck Institute and Roche. Eli Lilly has provided medication for a clinical trial led by SL as principal investigator. The other authors declare no competing interests.

Acknowledgments

This study was supported in part by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Grant Number 26670314 to TAF and by Great Britain Sasakawa Foundation Butterfield Award 4939 to ACi and TAF. ACi is supported by the NIHR Oxford Cognitive Health Clinical Research Facility and ACh by the International Short Visit Program of the Swiss National Foundation (Grant Number IZK0Z3_167541).

References

1 Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; **287:** 1840–47.

2 Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 2009; **166:** 42–49.

3 Khan A, Bhat A, Kolts R, Thase ME, Brown W. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? *CNS Neurosci Ther* 2010; **16:** 217–26.

4 Khin NA, Chen YF, Yang Y, Yang P, Laughren TP. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J Clin Psychiatry* 2011; **72:** 464–72.

5 Fountoulakis KN, Moller HJ. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *Int J Neuropsychopharmacol* 2011; **14:** 405–12.

6 Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG. Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* 2009; **118:** 1–8.

7 Montgomery SA. The failure of placebo-controlled studies. ECNP Consensus Meeting, September 13, 1997, Vienna. European College of Neuropsychopharmacology. *Eur Neuropsychopharmacol* 1999; **9:** 271–76.

8 Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013; **170:** 723–33.

9 Sinyor M, Levitt AJ, Cheung AH, et al. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry* 2010; **71:** 270–79.

10 Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol* 2009; **19:** 34–40.

11 Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. *Neuropsychopharmacology* 2003; **28:** 552–57.

12 Rutherford BR, Cooper TM, Persaud A, Brown PJ, Sneed JR, Roose SP. Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and dropout. *J Clin Psychiatry* 2013; **74:** 703–15.

13 Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics* 2000; **56:** 271–78.

14 Undurraga J, Baldessarini RJ. Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology* 2012; **37:** 851–64.

15 Furukawa TA, Salanti G, Atkinson L, et al. Comparative efficacy and acceptability of first- and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open* 2016; **6:** e010919.

16 Higgins JP, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. http://www.cochrane-handbook.org (accessed June 28, 2016).

17 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7:** 177–88.

18 Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol* 2005; **20:** 49–52.

19 Spineli LM, Higgins JP, Cipriani A, Leucht S, Salanti G. Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis. *Clin Trials* 2013; **10:** 378–88.

20 Carmody TJ, Rush AJ, Bernstein I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol* 2006; **16:** 601–11.

21 Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, Bandelow B. Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry* 2006; **67:** 1741–46.

22 Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970–2010. *Am J Psychiatry* 2013; **170:** 1335–44.

23 Nikolakopoulou A1, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health* 2014; **17:** 53–57.

24 Drazen JM. Sharing individual patient data from clinical trials. *N Engl J Med* 2015; **372:** 201–02.

25 Mancini M, Wade AG, Perugi G, Lenox-Smith A, Schacht A. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res* 2014; **51:** 21–29.

26 Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. *Lancet Psychiatry* 2015; **2:** 246–57.

27 Fava M. Implications of a biosignature study of the placebo response in major depressive disorder. *JAMA Psychiatry* 2015; **72:** 1073–74.

28 Cipriani A, Geddes JR. Placebo for depression: we need to improve the quality of scientific information but also reject too simplistic approaches or ideological nihilism. *BMC Med* 2014; **12:** 105.

29 Geddes JR, Cipriani A. Time to abandon placebo control in pivotal phase III trials? *World Psychiatry* 2015; **14:** 306–07.

30 Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013; **170:** 723–33.

31 Norman TC, Bountra C, Edwards AM, Yamamoto KR, Friend SH. Leveraging crowdsourcing to facilitate the discovery of new medicines. *Sci Transl Med* 2011; **3:** 88mr1.