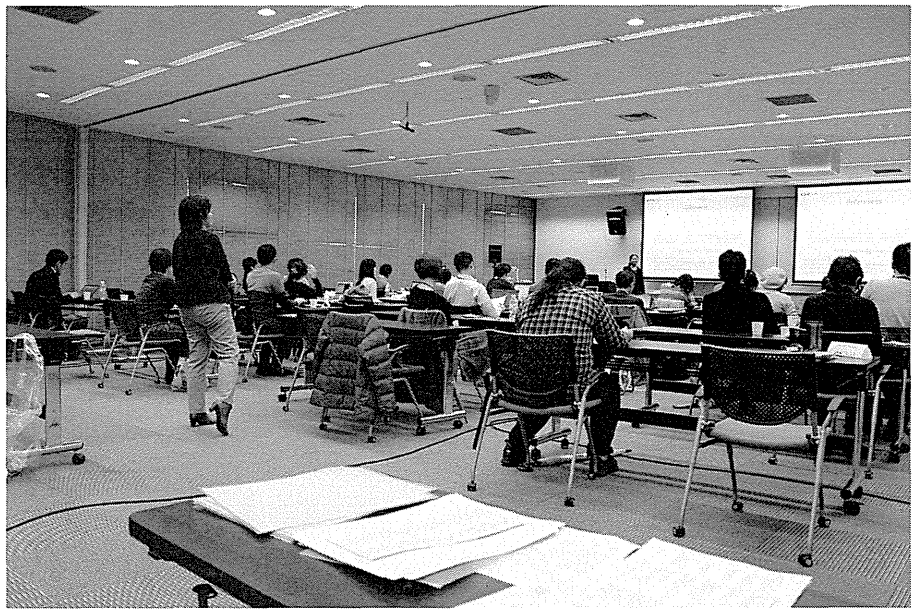
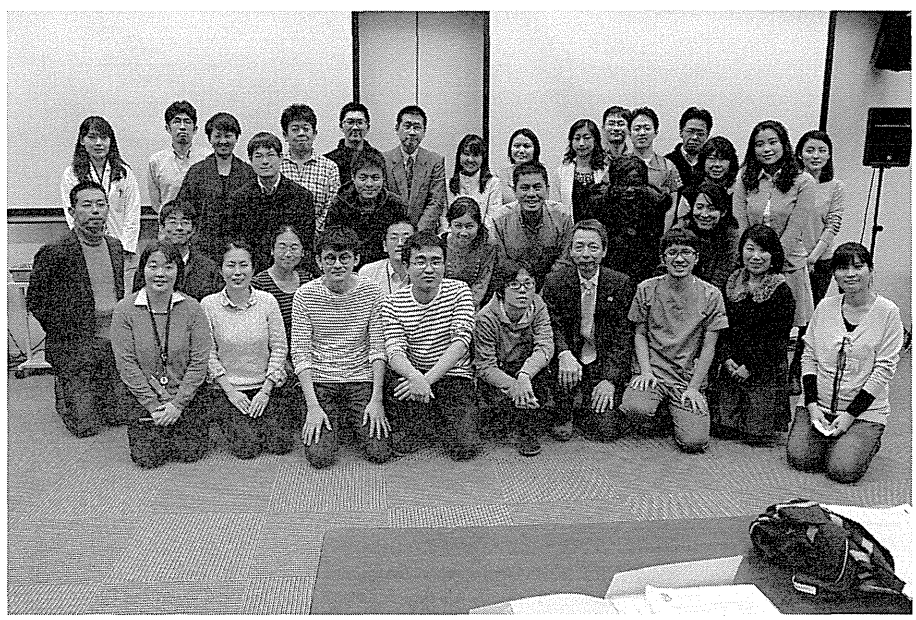




平成26年度厚生労働科学研究費補助金
(成育疾患克服等次世代育成基盤研究事業（健やか次世代育成総合研究事業）)
「母子保健に関する国際的動向及び情報発信に関する研究」分担研究報告書

ネットワークメタアナリシスの応用と批判的吟味ガイドライン

研究分担者 古川壽亮（京都大学大学院医学研究科　教授）

研究要旨

複数の治療選択肢を同時に比較するネットワークメタアナリシスが様々な臨床疑問に対して用いられるようになってきた。本研究では、さらに進んで、複数の治療選択肢の間の優劣の比較にとどまらず、ネットワークメタアナリシスの方法論的強みを生かした理論的研究を行った。第1の研究では、周産期うつ病や児童思春期の発達障害・情緒障害への効果が近年着目されている認知行動療法について、その特異性の有無を検討した。第2の研究では、精神療法の臨床試験において用いられる対照群の差異を明らかにした。さらに、このような実践を通じ、ネットワークメタアナリシのスの批判的吟味のガイドラインを作成した。

A. 研究目的

新世代のエビデンス統合方法として、ネットワークメタアナリシス network meta-analysis の発表が急激に増えている[1]。ネットワークメタアナリシスでは、複数の治療選択肢を比較した無作為割り付け比較試験(RCT)をすべて集めてきて、従来のメタアナリシスのように1対1で比較するのではなく、エビデンスのネットワーク全体を同時に統合する。すなわち、Aという治療選択肢とBという治療選択肢を比較するには、AとBとを直接に比較した直接比較に加えて、AとBとのあいだの効果の差は、Aとまた別の治療選択肢Cとの間の差、マイナス、BとCとの間の差と同じなるはずであるので、このような間接比較の結果も統合することができる。ネットワークの全体について直接比較と間接比較の統合を行うのがネットワークメタアナリシスである。

本研究では、ネットワークメタアナリシスの方法論を応用し、種々の治療選択肢の間での優劣の比較にとどまらず、理論的研究に応用する研究を行った。さらに、このような実践を通じ、ネットワークメタアナリシスの批判的吟味のガイドラインを作成した。

B. ネットワークメタアナリシス1

精神療法の効果は、その技法に特異的な部分と、すべての技法に共通の非特異的な部分に分けて考えることが出来る。従来の研究では、この特異的な効果というものが実際にあるのか否かが議論されてきた。

そこで本研究では、周産期うつ病や児童青年期の発達障害また情緒障害への効果が近年注目されている認知行動療法を取り上げ、ネットワークメタアナリシスの方法論を応用して、プラセボ精神療法を上回る要素があるか否かを検討した[2]。

B1. 方法

対象となる研究の種類：RCT、cross-over trial の場合はその前半のみ、cluster RCT を対象とする。

対象となる参加者：18歳以上75歳未満。

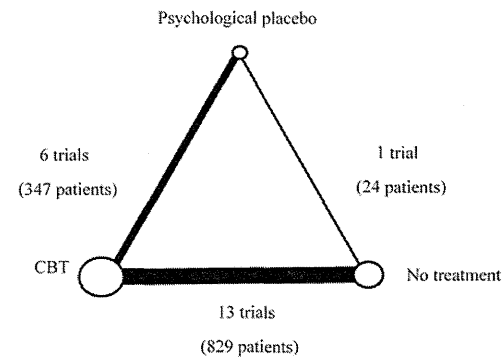
対象となる診断：操作的診断(DSM-IV, DSM-III-R, DSM-III, ICD-10, RDC, Feighner)による大うつ病急性期。確立された評価尺度の閾値によってエントリーされた場合も包含する。一方、治療抵抗性の大うつ病や、大うつ病の再発の予防を目的とした試験は除外する。
実験群介入：認知行動療法
対照群介入：無治療群、心理プラセボ
アウトカム尺度：主要アウトカムは、抑うつを測定する連続尺度に基づき判定された回復/改善とした。

研究の検索：Cochrane Collaboration Depression, Anxiety and Neurosis Group の CCDANCTR、引用文献リスト、個人的連絡を用いる
二人の独立した評価者が各研究が選択基準を満たしているかを検討し、二人の意見が不一致の場合は第三の著者と検討する。
データ抽出：二人の独立した評価者によってあらかじめ定められたデータを抽出する。二人の意見が不一致の場合は第三の著者と検討する。
欠損値の扱い：二値尺度については、脱落した者は不良なアウトカムであったと想定して ITT を行う。また連続尺度が与えられている場合は、Furukawa[3] により反応率を推定した。
報告バイアスの評価：出来る限りもれなく研究を同定すること、各研究において重要なアウトカムが欠落していないかを検討し場合によっては原著者に問い合わせる。十分な数の研究があれば漏斗図分析を行う。

治療効果の表現：二値尺度については OR を用いる
データの統合：ランダム効果モデルを使う
メタレグレッション：得られた結果の異質性の原因を検討するため
1) セッション数
2) 個人 vs 集団治療の別
3) ベースライン抑うつ重症度
4) 抗うつ剤の使用
についてメタレグレッションを行う。

（倫理面への配慮）
出版されたデータの二次利用であるので、倫理委員会の承認は要さない。

B2. 結果
同定された研究
われわれはこのプロトコルに則り、CCDANCTR などから 2011 年 6 月までの検索により該当 RCT の同定およびそれからのデータ抽出を進めた。
その結果、128 研究（3737 アーム）を同定した。このうち、認知行動療法、プラセボ精神療法、無治療を比較した RCT はすべてで 18 研究（1153 人）であった。
そのエビデンスネットワークは



の通りであった。包含された RCT の質は残念ながら全般に低かった。

ネットワークメタアナリシス
各比較の異質性は中等度でかつ出版バイアスも疑われなかったので、ネットワークメタアナリシスを行った。ネットワークのモデルフィットも十分と判定された。

	全体
CBT vs NT	2.24 (1.32 to 3.88)
CBT vs PP	1.30 (0.53 to 2.94)

PP vs NT	1.73 (0.67 to 4.84)
CBT specific component	35.0% (-99.5 % to 180.3%)

と、CBT vs 無治療は有意な結果であったが、CBT とプラセボ精神療法あるいは CBT と無治療の間には有意な差が認められなかった。CBT の特異的な効果は35%と推定されたが、その 95%信用区間はマイナスから 100%超と非常に広く、CBT の特異的な効果があるとは言いきれない結果であった。

メタレグレッション

そこで、アプリオリに設定された少数の変数についてメタ回帰を行ったところ、セッション数が有意な影響を示し、

	< 10sessions	**≥ 10sessions**
CBT vs NT	1.53 (1.02 to 2.28)	7.37 (3.74 to 15.15)
CBT vs PP	0.55 (0.27 to 1.20)	2.71 (1.42 to 5.33)
PP vs NT	2.72 (1.28 to 5.76)	2.72 (1.28 to 5.76)
CBT specific component	-159.6% (-958.4% to 90.6%)	50.4% (19.7% to 85.0%)

10 セッション以上の CBT は、プラセボ精神療法との間に有意な差を認めた。その場合、CBT の特異的な効果は、50.4% (19.7% to 85.0%)と推定された。

C. ネットワークメタアナリシス 2

薬物療法の臨床試験においてはプラセボピルが対照群に用いられ、平均への回帰、自然経過、Hawthorne 効果、そしてプラセボ効果をコントロールできるとされる。すなわち、実薬とプラセボピルとの間に見られた効果差は、実薬に特異的な薬理学的効果と推定される。

これにならって、精神療法の臨床試験においてもプラセボ精神療法が提案され、上記のネットワークメタアナリシス 1 で包含されたいくつかの研究のようにこれを用いた研究もあるが、厳密なプラセボ精神療法を考案することはしばしば困難である。そこで、精神療法の臨床試験では、無治療、通常治療、待機リストが対照群として用いられることがある。このうち、通常治療は何をもって通常治療とするかが地域と時代によって異なり統一的に評価することが難しいので、これは除外して考えるとしても、プラセボ精神療法、待機リスト、無治療のいずれを対照群に用いるかによって治療効果サイズの推定が異なってくるのではないかという疑念がかねてよりあった。

そこで本研究では、認知行動療法群、無治療対照群、待機リスト群、プラセボ精神療法群の 4 群間でネットワークメタアナリシスを行うことにより、対照群によって治療効果の推定が異なるか否かを検討した [4]。

C1. 方法

対象となる研究の種類：無作為割り付け比較試験(RCT)、cross-over trial の場合はその前半のみ、cluster RCT を対象とする。

対象となる参加者：18 歳以上 75 歳未満。

対象となる診断：操作的診断(DSM-IV, DSM-III-R, DSM-III, ICD-10, RDC, Feighner)による大うつ病急性期。確立された評価尺度の閾値によってエントリーされた場合も包含する。一方、治療抵抗性の大うつ病や、大うつ病の再発の予防を目的とした試験は除外する。

実験群介入：下記の認知的または行動的介入の 1 つ以上を用いた広義の認知行動療法 (CBT)

1) 認知再構成
2) 行動活性化
3) 問題解決
4) アサーション訓練
5) マインドフルネス

6) 除外される介入として、再発予防のための介入、治療者付きのセルフヘルプ、薬物との併用療法、夫婦療法、家族療法。

対照群介入：無治療(No treatment: NT)、待機(Waiting list: WL)、プラセボ精神療法(Placebo psychotherapy: PP)

アウトカム尺度：主要アウトカムは、抑うつを測定する連続尺度に基づき判定された回復/改善とした。

研究の検索：Cochrane Collaboration Depression, Anxiety and Neurosis Group の CCDANCTR、引用文献リスト、個人的連絡を用いる

二人の独立した評価者が各研究が選択基準を満たしているかを検討し、二人の意見が不一致の場合は第三の著者と検討する。

データ抽出：二人の独立した評価者によってあらかじめ定められたデータを抽出する。二人の意見が不一致の場合は第三の著者と検討する。

欠損値の扱い：二値尺度については、脱落した者は不良なアウトカムであったと想定して ITT を行う。また連続尺度が与えられている場合は、Furukawa[3] により反応率を推定した。

報告バイアスの評価：出来る限りもれなく研究を同定すること、各研究において重要なアウトカムが欠落していないかを検討し場合によっては原著者に問い合わせる。十分な数の研究があれば漏斗図分析を行う。

治療効果の表現：二値尺度については OR を用いる

データの統合：ランダム効果モデルを用いる

（倫理面への配慮）

出版されたデータの二次利用であるので、倫理委員会の承認は要さない。

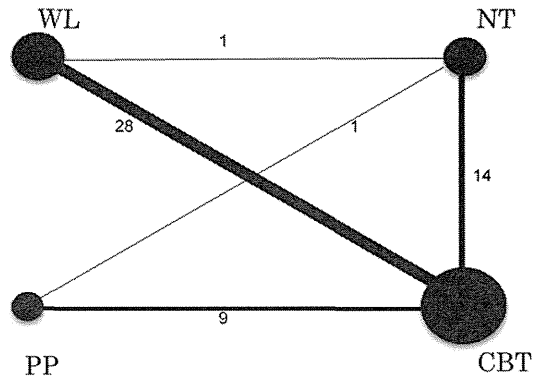
C2. 結果

同定された研究

われわれはこのプロトコルに則り、CCDANCTR などから 2011 年 6 月までの検索により該当 RCT の同定およびそれからのデータ抽出を進めた。

その結果、128 研究（3737 アーム）を同定した。このうち、認知行動療法、無治療、待機、プラセボ精神療法を比較した RCT はすべてで 49 研究（2730 人）であった。

そのエビデンスネットワークは



の通りであった。包含された RCT の質は残念ながら全般に低かった。

ペアワイズ　メタアナリシス

認知行動療法と核体症群との間のペアワイズのメタアナリシスの結果は、

	OR	p	I-squared	Egger's test p
CBT vs PP	1.60 (0.95 to 2.67)	0.077	18.8% (0% to 81.3%)	0.69
CBT vs NT	2.07 (1.35 to 3.18)	<0.001	38.8% (0% to 79.0%)	0.34
CBT vs WL	3.99 (2.76 to 5.77)	<0.001	31.5% (0% to 76.9%)	<0.001

で、CBT は無治療群および待機群よりも有意に優れていたが、プラセボ精神療法との差は有意ではなかった。

異質性を示す I-squared 値は中等度以下であった。また、待機群との比較には重大な出版バイアスが存在する可能性が示唆された。

ネットワークメタアナリシス

ネットワークのモデルフィットは十分と判定されたので、ネットワークメタアナリシスを行った。

	PP	NT	WL
CBT	1.65 (0.76, 3.13)	2.36 (1.31, 4.26)	6.26 (3.90, 10.1)
PP	—	1.62 (0.62, 3.59)	4.31 (1.74, 9.46)
NT	—	—	2.87 (1.33, 5.65)

と、認知行動療法はプラセボ精神療法との間では有意差がないが、無治療および待機群よりも有意に反応をもたらしやすかった。さらに、無治療は待機群よりも反応をもたらしやすいという結果であった。

しかし、出版バイアスを補正するネットワークメタアナリシスを行ったところ、

	PP	NT	WL
CBT	1.34 (0.70, 2.24)	1.63 (0.78, 3.15)	1.79 (0.97, 2.87)
PP	—	1.31 (0.54, 2.79)	1.44 (0.64, 2.78)
NT	—	—	1.22 (0.48, 2.42)

と、すべての有意差はなくなった。

D. ネットワークメタアナリシスの批判的吟味

われわれの研究グループは、これまでに薬物療法のネットワークメタアナリシス[5,6]、精神療法のネットワークメタアナリシス[7]、そして上記B、Cのような理論的応用を目指したネットワークメタアナリシス[2,4]を行ってきた。

これまでの経験に基づき、今回われわれは臨床家がネットワークメタアナリシスを臨床で利用するための批判的吟味のガイドラインを作成した[8,9]。このガイドラインは、メタアナリシスのユーザーズガイド[10]にならい、2部に分かれる。

第1部は、系統的レビューの過程の方法論的妥当性に関するチェックポイントで、通常のペアワイズ・メタアナリシスと同様である。

Users' Guides for Credibility of the Systematic Review Process

Did the review explicitly address a sensible clinical question?

Was the search for relevant studies exhaustive?

Were selection and assessments of studies reproducible?

Did the review present results that are ready for clinical application?

Did the review address confidence in effect estimates?

第2部は、得られた結果の質、つまり得られた結果の確実性に関するチェックポイントで、GRADEをネットワークメタアナリシスに適用したものである

Users' Guides for Certainty in the Results

Is there low risk of bias for each comparison?

Are there concerns about publication bias?

Are the results of individual studies in direct comparisons consistent?

Are the results of direct and indirect comparisons consistent?

Are confidence or credible intervals sufficiently narrow?

Are the treatment rankings trustworthy?

E. 結論

ひとつの臨床疑問に関して複数の治療を比較するネットワークメタアナリシスは今後、臨床判断にますます重要となってくるであろう。今回われわれはさらに進んで、ネットワークメタアナリシスを理論的研究に応用する研究を行った。また、ネットワークメタアナリシスを臨床家が適切に利用できるためのガイドラインを作成した。

引用文献

1. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann. Intern. Med.* 2013;159(2):130-137.
2. Honyashiki M, Furukawa TA, Noma H, et al. Specificity of CBT for depression: a contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research.* 2014;38:249-260.
3. 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(1):49-52.
4. Furukawa TA, Noma H, Caldwell DM, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr. Scand.* 2014;130(3):181-192.
5. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet.* 2009;373:746-758.
6. Miura T, Noma H, Furukawa TA, et al. Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: A network meta-analysis. *Lancet Psychiatry.* 2014;1(5):351-359.
7. Pompoli A, Furukawa TA, Imai H, Tajika A, Efthimiou O, Salanti G. Psychological therapies for panic disorder with or without agoraphobia in adults [Protocol]. *Cochrane Database Syst. Rev.* 2014(2):CD011004.
8. Foote CJ, Chaudhry H, Bhandari M, et al. Network meta-analysis: Users' Guide for Surgeons Part I - Credibility. *Clin. Orthod. Res.* in press.
9. Chaudhry H, Foote CJ, Guyatt G, et al. Network meta-analysis: Users' guide for surgeons Part II - Certainty. *Clin. Orthop.* in press.
10. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA.* 2014;312(2):171-179.

F. 研究発表

出版されたネットワークメタアナリシス

1) Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V & Churchill R (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. **Acta Psychiatrica Scandinavica**, 130, 181-192.
2) Honyashiki M, Furukawa TA, Noma H, Tanaka S, Chen P, Ichikawa K, Ono M, Churchill R, Hunot V & Caldwell DM (2014) Specificity of CBT for depression: a contribution from multiple treatments meta-analyses. **Cognitive Therapy and Research**, 38, 249-260.
3) Miura T, Noma H, Furukawa TA, Mitsuyasu H, Tanaka S, Stockton S, Salanti G, Motomura K, Shimano-Katsuki S, Leucht S, Cipriani

A, Geddes JR & Kanba S (2014) Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: A network meta-analysis. **Lancet Psychiatry**, 1, 351-359.

4) Miura T, Furukawa TA, Cipriani A, Motomura K, Mitsuyasu H, Tanaka S, Leucht S, Shimano-Katsuki S, Salanti G, Noma H, Stockton S, Geddes JR & Kanba S (2014) Bipolar treatment efficacy - Author's replay. **Lancet Psychiatry**, 1, 418-419.

情報発信に関する臨床疫学的研究

5) Tajika A, Ogawa Y, Takeshima N, Hayasaka Y & Furukawa TA (in press) Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. **British Journal of Psychiatry**.

6) Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y & Furukawa TA (2014) Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? **BMC Medical Research Methodology**, 14, 30.

7) Onishi A & Furukawa TA (2014) Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. **Journal of Clinical Epidemiology**, 67, 1320-1326.

8) Foote CJ, Chaudhry H, Bhandari M, Thabane L, Furukawa TA, Petrisor B & Guyatt G (in press) Network meta-analysis: Users' Guide for Surgeons Part I - Credibility. **Clinical Orthodontics and Research**.

9) Chaudhry H, Foote CJ, Guyatt G, Thabane L, Furukawa TA, Petrisor B & Bhandari M (in press) Network meta-analysis: Users' guide for surgeons Part II - Certainty. **Clinical Orthopaedics and Related Research**.

10) Cipriani A & Furukawa TA (2014) Advancing evidence-based practice to improve patient care. **Evidence-Based Mental Health**, 17, 1-2.

11) Furukawa TA (2014) How can we make the results of trials and their meta-analyses using continuous outcomes clinically interpretable? **Acta Psychiatrica Scandinavica**, 130, 321-323.

12) Furukawa TA (2014) Case study in psychiatry. In **Network Meta-Analysis: Evidence Synthesis with Mixed Treatment Comparison**. (ed Biondi-Zoccai G). Hauppauge, NY: Nova Science Publishers.

13) Furukawa TA, Strauss S, Bucher HC, Agoritsas T & Guyatt G (2014) Diagnostic tests. In **Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice** (3rd edn). (eds Guyatt G, Drummond R, Meade MO & Cook DJ), pp. 345-358. New York: The McGraw-Hill Companies, Inc.

14) Furukawa TA, Scott I & Guyatt G (2014) Measuring patients' experience. In **Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice** (3rd edn). (eds Guyatt G, Drummond R, Meade MO & Cook DJ), pp. 219-234. New York: The McGraw-Hill Companies, Inc.

15) Furukawa TA & Guyatt G (2014) An illustration of bias and random error. In **Users' Guides to the Medical**

Literature: A Manual for Evidence-Based Clinical Practice (3rd edn). (eds Guyatt G, Rennie D, Meade MO & Cook DJ), pp. 103-106. New York: The McGraw-Hill Companies, Inc.

コクランレビュー

16) Imai H, Tajika A, Chen P, Pompoli A, Guaiana G, Castellazzi M, Bighelli I, Girlanda F, Barbui C, Koesters M, Cipriani A & Furukawa TA (2014) Azapirone versus placebo for panic disorder in adults. **Cochrane Database of Systematic Reviews**, CD010828.
17) Purgato M, Papola D, Gastaldon C, Trespidi C, Magni LR, Rizzo C, Furukawa TA, Watanabe N, Cipriani A & Barbui C (2014) Paroxetine versus other anti-depressive agents for depression. **Cochrane Database of Systematic Reviews**, CD006531.

コクランレビュープロトコル

18) Imai H, Tajika A, Chen L, Pompoli A & Furukawa TA (2014) Psychological therapies versus pharmacological interventions for panic disorder [Protocol]. **Cochrane Database of Systematic Reviews**, CD011170.
19) Kimachi M, Furukawa TA, Kimachi K, Goto Y & Fukuhara S (2014) New oral anticoagulants versus warfarin for preventing stroke and systemic embolic events among atrial fibrillation patients with chronic kidney disease [Protocol]. **Cochrane Database of Systematic Reviews**.
20) Katsura M, Kuriyama A, Takeshima T, Fukuhara S & Furukawa TA (2014) Preoperative inspiratory muscle training for postoperative pulmonary complications in adult patients undergoing cardiac and major abdominal surgery [Protocol]. **Cochrane Database of Systematic Reviews**.
21) Onishi A, Sato A, Iwasaku M & Furukawa TA (2014) Human parathyroid hormone for treating glucocorticoid-induced osteoporosis [Protocol]. **Cochrane Database of Systematic Reviews**.
22) Pompoli A, Furukawa TA, Imai H, Tajika A, Efthimiou O & Salanti G (2014) Psychological therapies for panic disorder with or without agoraphobia in adults [Protocol]. **Cochrane Database of Systematic Reviews**, CD011004.
23) Sakai K, Fujita K, Sozu T, Nakayama T & Furukawa TA (2015) Eradication of helicobacter pylori for iron metabolism [Protocol]. **Cochrane Database of Systematic Reviews**, CD011480.
24) Takeshima N, Furukawa TA, Hayasaka Y, Ogawa Y & Tajika A (2014) Ramelteon for primary insomnia [Protocol]. **Cochrane Database of Systematic Reviews**, CD011049.

G. 知的財産権の出願・登録状況

1. 特許取得　なし
2. 実用新案登録　なし
3. その他　なし

Acta Psychiatrica Scandinavica

Acta Psychiatr Scand 2014: 130: 181–192
All rights reserved
DOI: 10.1111/acps.12275

ACTA PSYCHIATRICA SCANDINAVICA

Meta-analysis

Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis

Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis.

Objective: Various control conditions have been employed in psychotherapy trials, but there is growing suspicion that they may lead to different effect size estimates. The present study aims to examine the differences among control conditions including waiting list (WL), no treatment (NT) and psychological *placebo* (PP).
Method: We comprehensively searched for all randomized controlled trials (RCTs) comparing cognitive-behaviour therapies (CBT) against various control conditions in the acute phase treatment of depression, and applied network meta-analysis (NMA) to combine all direct and indirect comparisons among the treatment and control arms.
Results: We identified 49 RCTs (2730 participants) comparing WL, NT, PP and CBT. This network of evidence was consistent, and the effect size estimates for CBT were substantively different depending on the control condition. The odds ratio of response for NT over WL was statistically significant at 2.9 (95% CI: 1.3–5.7). However, the quality of evidence, including publication bias, was less than ideal and none of the preplanned sensitivity analyses limiting to high-quality studies could be conducted, while findings of significant differences did not persist in *post hoc* sensitivity analyses trying to adjust for publication bias.
Conclusion: There may be important differences in control conditions currently used in psychotherapy trials.

T. A. Furukawa[1], H. Noma[2], D. M. Caldwell[3], M. Honyashiki[4], K. Shinohara[4], H. Imai[4], P. Chen[4], V. Hunot[5], R. Churchill[5]

[1]Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, [2]Department of Data Science, Institute of Statistical Mathematics, Tokyo, Japan, [3]School of Social and Community Medicine, University of Bristol, Bristol, UK, [4]Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan and [5]Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, Bristol, UK

Key words: waiting lists; *placebo*; control groups; cognitive therapy; clinical trials

Toshi A. Furukawa, 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. E-mail: furukawa@kuhp.kyoto-u.ac.jp

Accepted for publication March 12, 2014

Summations

- We identified a dense, well-connected, homogeneous and consistent network of evidence connecting cognitive-behaviour therapies (CBT), psychological *placebo* (PP), no treatment (NT) and waiting list (WL) in the acute phase treatment of adult depression.
- Applying the network meta-analysis (NMA), the effect size estimates for CBT were substantively different, depending on which control condition it was compared against.
- The indirect comparison between NT and WL revealed that the former was significantly superior to the latter in producing response (OR = 2.9, 95% confidence interval: 1.3–5.7).

Considerations

- The quality of randomized evidence constituting the above network was less than ideal.
- Statistically significant differences in the primary analyses were lost when we applied exploratory sensitivity analyses using network meta-regression to adjust for publication bias.

Introduction

The need for a control condition in evaluating efficacy/effectiveness of any health intervention is accepted beyond doubt but what constitutes an appropriate control in psychotherapy trials has long been hotly debated (1, 2).

The level of control required is dependent on what factors one wishes to control in trials. Maximally, we would like to control for the regression towards the mean, the natural course of the disease, the Hawthorne effect (the effect of being observed and evaluated), and the *placebo* effect (the effect of believing to be treated for benefit). In pharmacotherapy trials, the pill *placebo* control would control for all these factors, and it therefore follows that, if a certain chemical compound turns out to be superior to the pill *placebo* in a comparative trial that compound will be believed to have a specific efficacy for the disorder. Following this model, earlier theorists have advocated for a psychological *placebo* (PP), which would similarly control for the four factors above in psychotherapy trials (3). In reality, such has not been easy to implement and many psychotherapy trials have relied on less than ideal control conditions, including treatment as usual (TAU), no treatment (NT) and waiting list (WL). Many systematic reviews have traditionally lumped all these into one control condition in evaluating psychotherapies (4–6) or various human interventions (7). TAU, sometimes also referred to as usual care (UC), is simply too heterogeneous across studies and countries to be regarded one entity in a review (8). We therefore did not include the so-called TAU or UC conditions in the present review. However, even aside from them, there are now growing theoretical and empirical concerns that different control conditions may lead to different effect size estimates (9–12).

When several alternative interventions or controls are compared in a series of randomized trials, we now have an advanced method of evidence synthesis called network meta-analysis (NMA). Traditional, head-to-head meta-analyses can compare only two interventions at a time. NMA allows assessment of relative effectiveness of all the included interventions by integrating data from both direct and indirect comparisons, thus borrowing strength from the entire network of randomized trials (13, 14). NMA is therefore particularly suitable for the assessment of differences among alternative control conditions, each of which is compared against the active treatment but only a few of which are compared directly among themselves.

Aim of the study

The present study aims to differentially examine no treatment, waiting list and psychological *placebo* control conditions by applying network meta-analysis to the relevant body of randomized controlled trials of cognitive-behaviour therapies for adult depression.

Material and methods

Study eligibility

The eligibility criteria for the studies were as follows.

Study design. All relevant randomized controlled studies. Studies in which two relevant conditions were compared and in which a cointervention (except for protocolized antidepressant treatment intended as combination treatment) was simultaneously provided were accepted when the cointervention was equally administered in both arms. Quasi-randomized controlled trials, in which treatment assignment was decided through methods such as alternate days of the week, were excluded.

Participants. Patients between ages 18 and 75, of both sexes, with a primary diagnosis of acute phase unipolar depression,

i) Diagnosed as such according to any of the following operationalized criteria: Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV or ICD-10 or
ii) Identified as such using a validated clinician or self-report depression symptom questionnaire, based on a recognized threshold.

Because differences in cognitive capabilities among younger or elderly people may influence the effectiveness and acceptability of psychotherapies (11, 15), we excluded studies of children and adolescents aged ≤18 or of older people where the mean age of participants was ≥75.

Studies focusing on patients with chronic depression or treatment-resistant depression were excluded. Studies that focused on depression among participants who all had a concurrent primary diagnosis of another mental or physical disorder were excluded. Existence of concurrent secondary diagnosis of another disorder was allowed.

Interventions. The control conditions of interest in this study included the following:

i) PP: a control condition that was regarded as inactive by the researchers but was to be

perceived as active by the participants. Furthermore, the number and duration of the sessions as well as the qualification of the therapists had to be equivalent with those of the active treatment in the same study.

ii) NT: a control condition in which the participants receive no active treatment during the study and in which they do not expect to receive such after the study is over.

iii) WL: a control condition in which the participants receive no active treatment during the study but are forewarned that they can receive one after the study period is over.

In both NT and WL, participants may or may not receive some medical care during the course of the study on a naturalistic basis.

The active intervention chosen to be contrasted in this study was CBT, because it is by far the best studied school of psychotherapy for depression (16) and was therefore likely to form the densest network in differentiating various control conditions. In this study, CBT was broadly conceived as comprising one or more of the following cognitive and behavioural components.

i) Cognitive restructuring, which aims at monitoring and modifying the patient's dysfunctional beliefs that lead to depressed feelings.

ii) Behavioural activation, which aims at helping the patient increase his/her frequency and quality of pleasant activities and/or goal-directed activities.

iii) Problem solving, which teaches the patient a staged and structured approach to pragmatic solution of his/her problems.

iv) Assertion training, which helps the patient initiate and negotiate interpersonal relationships by teaching him/her the skill to express honest feelings and requests.

v) Acceptance and mindfulness, which places less emphasis on rational challenging of thoughts but aims to transform the relationship between the experience of symptoms and difficult thoughts/feelings.

CBT had to be delivered through face-to-face meetings between the patient and therapist. Both individual and group formats were accepted.

Outcomes. Our primary outcome was the number of patients who responded to treatment, based on changes on Hamilton Rating Scale for Depression (HAMD) (17), Beck Depression Inventory (BDI) (18) or any other validated depression scale at the end of the acute phase treatment. Acute treatment was defined as an 8-week treatment in the analyses. If 8-week data were not available, we used data ranging between 4–16 weeks and the time point given in the original study as the study endpoint was given the preference. Many studies defined response by 50% or greater reduction on the rating scale, but some studies defined it using Jacobson's Reliable Change Index (19). In the latter case, we adopted the original study authors' definitions. If the original authors reported several outcomes, observer-rated scales were preferred to self-reported scales; among the former, HAMD and among the latter, BDI was given preference.

Intention-to-treat analyses were based on the total number of randomly assigned participants, irrespective of how the original study investigators analysed the data, by assuming all drop-outs to be non-responders. For studies in which the exact numbers of participants who had responded were not reported, but means and standard deviations for continuous depression scales were reported, the number of responders was calculated by using a validated imputation method (20, 21).

In a few instances where we were unable to obtain the number of participants randomized at baseline (e.g. the number of total drop-outs after initial randomization is reported but no separate number of drop-outs on each treatment arm is provided nor obtainable from the original authors), we still included such studies using the analysed numbers of participants.

For three- or more-armed trials in which a control condition was compared against two or more variants of CBT (e.g. cognitive therapy and problem solving), all CBT arms were collapsed into one group.

Data collection

To identify the relevant studies, we searched two clinical trials registries created and maintained by the Cochrane Depression, Anxiety and Neurosis Group (CCDAN), the CCDANCTR-Studies and CCDANCTR-References. Reports of trials for inclusion in the Group's registers are collated from weekly, generic searches of MEDLINE (1950–), EMBASE (1974–) and PSYCINFO (1967–) and quarterly searches of the Cochrane Central Register of Controlled Trials. Reports of trials are also sourced from the World Health Organization's trials portal (ICTRP), clinicaltrials.gov, drug companies' websites, the hand searching of key journals, dissertation abstracts, conference proceedings and other non-Cochrane systematic reviews and meta-analyses. Details of CCDAN's generic search strategies can be found in the Cochrane Collaboration Depression, Anxiety and Neurosis

Group's webpage (http://ccdan.cochrane.org/). The CCDANCTR-Studies register was searched using the following terms: Condition = (depress* or dysthymi*) and Intervention = (*therap* or training). This search was supplemented by corresponding searches in CCDAN-References register, CINAHL and PSYNDEX. The additional search strategies for the databases other than the Studies register can be found in Shinohara et al. (22). The reference lists of all selected studies were searched for additional published reports and citations to unpublished studies. Relevant review papers were checked. The most recent updated search for this review was done in February 2012.

Two review authors (RC, VH) examined the abstracts of all publications obtained through the search strategy. Full articles of all the studies identified by any of these review authors, were obtained. Conflicts of opinion regarding eligibility of a study were discussed with a third review author, having retrieved the full paper and consulted the authors if necessary, until consensus was reached. External subject or methodological experts were consulted if necessary.

Data from each study were extracted independently by at least three review authors. Any disagreement was discussed with an additional review author and where necessary, the original study authors were contacted for further information. Information relating to study population, interventions, comparators, potential biases in the conduct of the trial and outcomes were abstracted from the original reports into specially designed paper forms then double-entered into a spreadsheet.

Assessment of risk of bias

Two independent review authors assessed the risk of bias in the selected studies, using the Cochrane Collaboration 'risk of bias' tool (23). The following seven domains were considered:

i) Sequence generation
ii) Allocation concealment
iii) Blinding of therapist
iv) Blinding of participant
v) Blinding of assessor for primary outcome
vi) Incomplete outcome data
vii) Selective outcome reporting

In addition, the following risks of bias specific to psychotherapy trials were systematically appraised.

viii) Researcher allegiance/Conflict of interest: did the researcher(s) have vested interest for or against the therapies under examination?
ix) Treatment fidelity: was the therapy monitored against a manual or a scale through audio- or videotapes?
x) Therapist qualification: were the therapists qualified to deliver psychotherapy and have they had specialist training for the intervention they are providing?
xi) Therapist allegiance/conflict of interest: did the therapists have vested interest for or against the therapies they were providing?
xii) Other sources of bias: was the study apparently free of other problems that could put it at a high risk of bias?

Any disagreement was discussed with a third review author. Where necessary, the authors of the studies were contacted for further information.

Analyses

Individual trials constitute pairwise, head-to-head meta-analyses, which then constitute the network for the NMA. In the following, we therefore first examined the robustness of each pairwise meta-analysis that forms a part of the evidence network in this study. We then proceeded to examine the robustness of the entire NMA. Given the potential clinical heterogeneity of the populations studied as well as the CBT approaches used in the studies, we used the random-effects model in all analyses.

Pairwise meta-analyses. We conducted pairwise meta-analyses by synthesizing studies that compared the same intervention/control condition with a random-effects model to incorporate the assumption that different studies assessed different, yet related, treatment effects (24). We examined between-studies heterogeneity through visual inspection of the forest plots and also by the I^2 statistics (23). We followed the following rough guide to interpretation of I^2 statistics:

0–40%: might not be important;
30–60%: may represent moderate heterogeneity;
50–90%: may represent substantial heterogeneity;
75–100%: considerable heterogeneity.

We also drew funnel plots and used Egger's tests (25) to assess small study effects, including possible publication bias. We conducted these analyses using R version 2.15.2 and metafor version 1.6 (26).

Network meta-analyses. We implemented random-effects network meta-analyses, taking into account the heterogeneity of treatment effects across studies in the Bayesian framework using OPENBUGS 3.2.1. We modelled the binary outcomes in every

treatment group of every study using the logistic regression model, incorporating the heterogeneity across studies by random effects. We evaluated the comparative response rates of CBT and the three control conditions. We assessed the Bayesian estimates and 95% credible intervals (CIs) of the odds ratios, and evaluated significance using the CIs (according to whether the CI included the null values).

A key assumption of the NMA model is the consistency of the network, that is, that direct and indirect evidence on the same comparisons does not disagree beyond random errors. For all triangles contained in the network, the difference between the direct and indirect estimates was examined by Bucher's test of inconsistency (27). We also evaluated model fit of the consistent and inconsistent models using the residual deviance statistics and deviance information criterion (DIC) (28). In a well-fitting model, the residual deviance should be close to the number of data points, and the model fit should be better for the consistency model than for the inconsistency one in terms of the residual deviance as well as DIC.

Sensitivity analyses. We *a priori* had selected allocation concealment, assessor blinding, treatment fidelity and imputation of numbers of responders as potentially important effect modifiers to be examined in sensitivity analyses to limit the included studies to those at low risk of bias. We conducted additional meta-regression analyses using random-effects network meta-regression models (29) to examine potential effect moderators such as the mean age of participants, the type of rating scales (clinician-rated vs. self-rated), publication status (published vs. dissertation) and therapy format (individual vs. group). The effect of assuming all drop-outs to be non-responders was examined by the completers analysis. In addition, leave-one-out analyses (i.e. omitting one study at a time) were performed to evaluate the influence of individual studies to the overall NMA results.

When substantive small study effects were noted, we conducted meta-regression analyses to evaluate their influences involving the study-specific variances as a covariate (30).

Results

Study selection

Of 6710 studies identified through electronic search and reference search, 186 full text articles were retrieved, of which 57 studies satisfied the eligibility criteria for the present study (Fig. 1). Of these, we were unable to include eight studies in the present study, as there was critical information lacking in the original reports and unavailable from the original authors upon request (31–38).

Thus, we were able to include 49 randomized studies, representing 117 treatment arms and involving 2730 participants. Figure 2 shows the final evidence network. There were nine studies comparing CBT against PP, 14 studies comparing CBT against NT, and 28 studies comparing CBT against WL. One study compared CBT, PP and NT (39), and another compared CBT, PP and WL (40), thus contributing one study each to the comparison PP vs. NT and to the comparison PP vs. WL, respectively.

The characteristics of the included studies are listed in Table 1. The mean drop-out rates were 19.5% on CBT, 25.5% on PP, 35.0% on NT and 10.8% on WL. The risks of bias of all the included studies are illustrated in Fig. 3.

Pairwise meta-analyses

Table 2 summarizes the results of pairwise meta-analyses. In comparison with the comparator arm CBT, PP tended to be inferior (OR = 1.60, 95% CI: 0.95–2.67), NT was definitely inferior (2.07, 1.35–3.18) and WL gravely inferior (OR = 3.99, 2.76–5.77). Two single studies comparing NT and WL against the common comparator PP were in line with the above pairwise meta-analyses, showing a larger effect size for PP against WL than against NT.

Though not definitive, because the 95% CI were wide, the I^2 statistics hinted at low to moderate heterogeneity for all these pairwise meta-analyses. The funnel plot analysis suggested strong small study effects in the comparison CBT vs. WL but not in the comparisons CBT vs. NT or CBT vs. PP.

Network meta-analysis

The baseline characteristics of included studies were similarly distributed across the comparisons and consistency was therefore likely to hold across the network. There were two closed loops in our network (Fig. 2), neither of which was suggestive of inconsistency beyond chance according to Bucher's test ($P = 0.60$ for CBT-PP-NT, and $P = 0.49$ for CBT-PP-WL).

There was little to choose between the inconsistency and consistency models both in terms of residual deviance and DIC (Residual deviance and DIC were 108.5 and 181.3 for consistency model and 108.4 and 181.3 for inconsistency model,

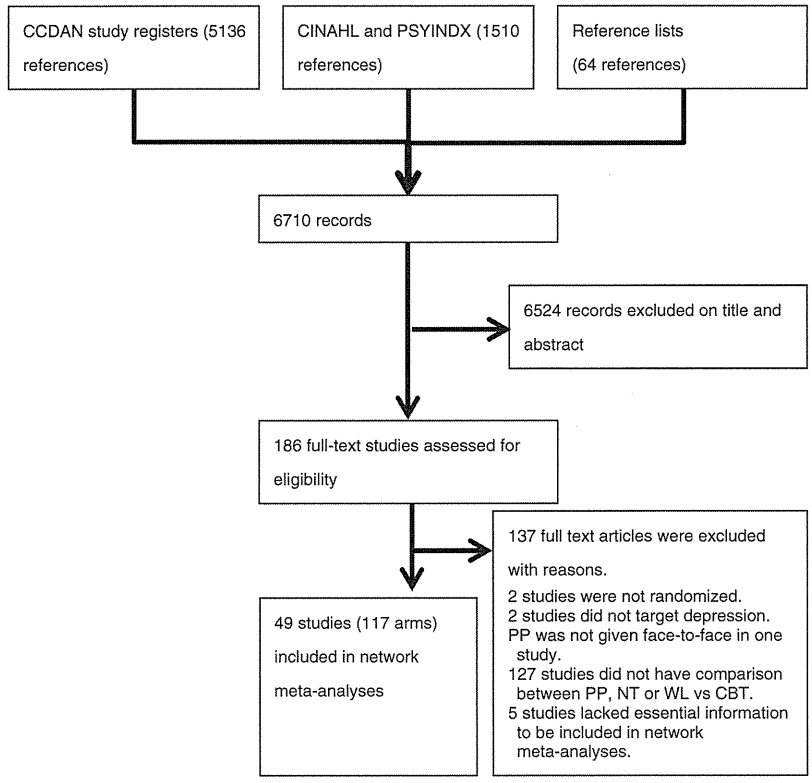


Fig. 1. PRISMA flowchart for the identification of relevant trials.

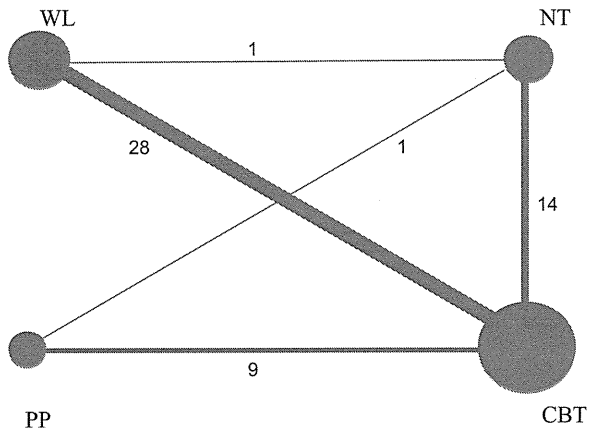


Fig. 2. Evidence network constituted by the 52 included comparisons.

respectively, for 98 data points). Owing to the way in which the residual deviance is calculated, zero cells (in which the number of responders is zero), can cause computational difficulties. We therefore further explored model fit by (i) excluding trials with zero cells and (ii) by applying a continuity correction of 0.5 to 0 cells. Given the reasonable concordance between the residual deviance and the number of data points when we excluded zero cells (83.0 for 82 data points) and applied continuity correction (102.1 for 98 data points), in conjunction with the results of Bucher's test above, the consistency model was judged to have satisfactory model fit. These results are shown in Table 3.

Sensitivity analyses

We were unable to conduct any of the preplanned sensitivity analyses, as there were only up to a quarter of studies at low risk of bias for each of them (14 for allocation concealment, 7 for assessor blinding, 12 for treatment fidelity and 4 for imputation). None of the regression coefficients of the meta-regression examining possible effect moderators turned out to be statistically significant [−0.024 (95% CI: −0.056 to 0.006) for age, −0.899 (−1.843 to 0.024) for rating scale, −0.442 (−1.399 to 0.520) for publication status, and 0.004 (−0.798 to 0.762) for therapy format].

The results of the NMA based on the completers only were essentially similar to our primary results. For example, the odds ratio of response for NT over WL was statistically significant at 2.6 (95% CI: 1.2–5.9). Also, all of the leave-one-out estimates were distributed around the overall NMA estimates with narrow ranges, suggesting that the

Table 1. Characteristics of the included studies

Study	No of participants	Included disorders	Baseline BDI	Types of CBT	Control conditions	Format	No of sessions	Concomitant pharmaco-therapy	Outcome scale
Arean (1993) (44)	48	MDD	22.5	PS	WL	G	12	Not allowed	HAMD
Ayen (2004) (45)	22	MDD+	22.0	Other CBT	WL	G	12	Allowed	BDI
Barrera (1979) (46)	20	Depression	25.1	BA	WL	G	5	Unclear	MMPI-D
Besyner (1979) (47)	20	Depression	24.9	CT	PP	G	4	Unclear	BDI
Brown (1984) (48)	63*	MDD+	31.8	CWD	WL	G/I	12	Allowed	BDI
Carrington (1979) (49)	20*	Depression	26.9	CT	WL	I	12	Unclear	BDI
Collins (1996) (50)	89	MDD	23.4	CT	WL	G	12	Allowed	BDI
Dowrick (Rural Finland) (1996)† (51)	50	MDD+	21.1	PS	NT	I	6	Allowed	BDI
Dowrick (Urban Finland) (1996)† (51)	47	MDD+	21.3	PS	NT	I	6	Allowed	BDI
Dowrick (Ireland, Urban&Rural) (1996)† (51)	38	MDD+	23	CWD	NT	G	8	Allowed	BDI
Dowrick (Rural Norway) (1996)† (51)	61	MDD+	19.2	CWD	NT	G	8	Allowed	BDI
Dowrick (Urban Norway) (1996)† (51)	67	MDD+	21	CWD	NT	G	8	Allowed	BDI
Dowrick (Urban Spain) (1996)† (51)	30	MDD+	22	PS	NT	I	6	Allowed	BDI
Dowrick (Rural UK) (1996)† (51)	49	MDD+	26	PS	NT	I	6	Allowed	BDI
Dowrick (Urban UK) (1996)† (51)	84	MDD+	24.8	PS, CWD	NT	I/G	6/8	Allowed	BDI
Embling (2002) (52)	38	MDD	31.0	CT	WL	G	12	Allowed	BDI
Epstein (1987) (53)	22	MDD+	25.8	CT	WL	G	8	Not allowed	BDI
Faramarzi (2008) (54)	82	Depression	19.9	CT	NT	G	10	Not allowed	BDI
Fuchs (1977) (40)	28*	Depression	22.8	Other CBT	PP, WL	G	6	Unclear	BDI
Hamamci (2006) (55)	24	Depression	28.4	CT	NT	G	11	Not allowed	BDI
Hamdan-Mansour (2009) (56)	84	Depression	24.1	Other CBT	NT	G	10	Unclear	BDI
Hautzinger (2004) (57)	100	MDD+	NA	CT	WL	G	12	Allowed	HSCL-D
Hayman (1980) (58)	28	Depression	18.5	AT	WL	G	8	Not allowed	BDI
Hegerl (2010) (59)	120	MDD+	NA	CT	PP	G	10	Allowed	HAMD
Hess-Homeier (1981) (60)	14	Depression	24.5	CT	WL	I	24	Not allowed	BDI
Kelly (1982) (61)	24	MDD	25.1	BA, REBT	PP	G	6	Allowed	BDI
Malouf (1984) (62)	53	Depression	20.6	PS, REBT	WL	G	4	Unclear	BDI
Miranda (2003) (63)	179	MDD	NA	CT	NT	I/G	8	Not allowed	HAMD
Nezu (1986) (64)	21	MDD	22.5	PS	WL	G	8	Not allowed	BDI
Nezu (1989) (65)	43	MDD	27.0	PS	WL	G	10	Not allowed	HAMD
Pace (1977) (66)	16	Depression	22.3	BA	PP	I	7	Unclear	BDI
Pace (1993) (67)	99	Depression	17.0	CT	WL	I	7	Not allowed	BDI
Pecheur (1980) (68)	21	MDD	22.7	CT	WL	G	8	Not allowed	HAMD
Pellowe (2006) (69)	54	Depression	13.8	ACT	PP	G	4	Allowed	BDI
Propst (1980) (39)	47	Depression	15.4	Other CBT	PP, NT	G	8	Not allowed	BDI
Propst (1992) (70)	49*	MDD+	17.3	CT, other CBT	WL	I	18	Not allowed	HAMD
Ross (1985) (71)	67	MDD	27.8	CT	WL	I/G	12	Allowed	BDI
Schmidt (1983) (72)	44	Depression	24.9	Other CBT	WL	I/G	8	Unclear	BDI
Schmitt (1988) (73)	40	MDD	27.1	PS, AT	WL	G	12	Unclear	HAMD
Serfaty (2009) (74)	137	MDD+	26.8	CT	PP	I	12	Allowed	BDI
Taylor (1977) (75)	28	Depression	21.2	CT, BA, other CBT	WL	I	6	Not allowed	BDI
Usaf (1990) (76)	60	MDD	27.5	CWD	WL	G	10	Not allowed	BDI
Wilson (1982) (77)	21*	Depression	26.0	BA	PP	I	7	Unclear	BDI
Wilson (1983) (78)	25	Depression	24.0	BA, CT	WL	I	8	Not allowed	HAMD
Wollersheim (1991) (79)	16	MDD	25.8	CBT other	WL	G	10	Unclear	BDI
Wong (2008a) (80)	337	MDD	20	CT	WL	G	10	Allowed	BDI
Wong (2008b) (81)	96	MDD	23.9	CT	WL	G	10	Allowed	BDI
Wright (2000) (82)	45	MDD	29.7	CT, CBT other	WL	I	9	Not allowed	HAMD

PP, psychological placebo; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapy; MDD, major depressive disorder diagnosed by operationalized diagnostic criteria; MDD+, major depressive disorder and depressive disorder NOS as diagnosed; ACT, acceptance and commitment therapy, AT, assertion training; BA, behavioural activation; CT, cognitive therapy; CWD, coping with depression course; PS, problem solving; REBT, rational emotive-behaviour therapy; G, group; I, individual; BDI, beck depression inventory; HAMD, Hamilton Rating Scale for depression; HSCL-D, Hopkins Symptom Checklist Depression Scale; MMPI-D, Minnesota Multiphasic Personality Inventory Depression Scale by operationalized diagnostic criteria.

*For these studies, randomized N was not available. Instead, we used number of participants analysed.

†Dowrick (1996) ENREF_41 reports nine independently conducted, albeit according to concerted protocols, RCTs. Two of these RCTs conducted in Ireland were reported in an amalgamated form in the definitive report (83) and is therefore treated as one trial in this meta-analysis.

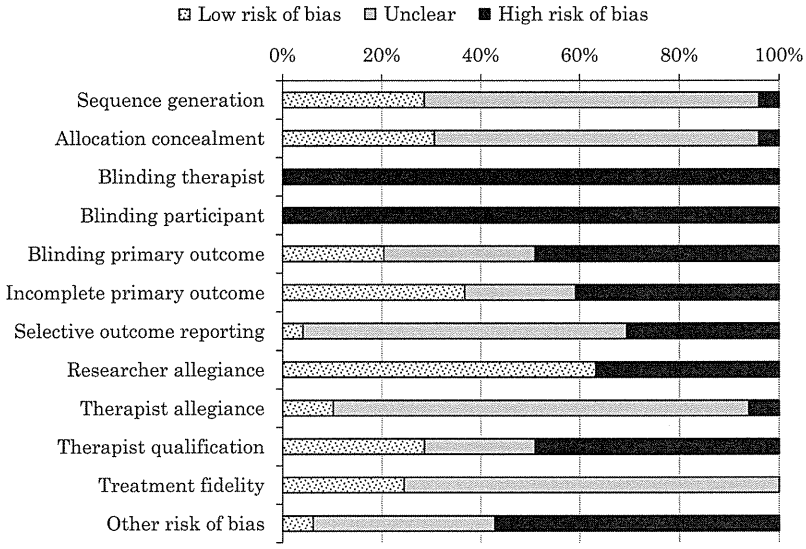


Fig. 3. Risks of bias of the 52 included trials.

Table 2. Results of pairwise meta-analyses

	Number of studies	Number of participants	OR	*P*	I^2, % (%)	Egger's test *P*
CBT vs. PP	9	446	1.60 (0.95–2.67)	0.077	18.8 (0–81.3)	0.69
CBT vs. NT	14	829	2.07 (1.35–3.18)	<0.001	38.8 (0–79.0)	0.34
CBT vs. WL	28	1486	3.99 (2.76–5.77)	<0.001	31.5 (0–76.9)	<0.001
PP vs. NT	1	24	2.04 (0.40–10.56)	0.394	–	–
PP vs. WL	1	20	6.00 (0.53–67.7)	0.147	–	–

CBT, cognitive-behaviour therapies; PP, psychological *placebo*; WL, waiting list.

Table 3. Results of network meta-analysis

	PP	NT	WL
CBT	1.65 (0.76–3.13)	2.36 (1.31–4.26)	6.26 (3.90–10.1)
PP	–	1.62 (0.62–3.59)	4.31 (1.74–9.46)
NT	–	–	2.87 (1.33–5.65)

OR shows the effectiveness of the intervention on the left over that on the top. 95% credible intervals in parentheses. Resdev = 109.2 (Data points: 98), DIC = 181.2.
PP, psychological *placebo*; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapies.

overall NMA estimates obtained in this study are stable and no influential outliers were involved.

Given the strong small study effects in the comparison between CBT and WL, we ran two additional meta-regressions incorporating the assumption that there is such a bias favoring CBT over WL (30). Two models were hypothesized: The first model assumed that there is such a bias not only between CBT and WL but that there are other similar (exchangeable) levels of biases in all the other comparisons in the network; the second model, on the other hand, assumed that there is no small study effects between CBT and PP but that there are exchangeable small study effects biases in all the other comparisons. The DIC suggested that the second model was a better fitting model. In this model, the regression coefficient representing small study effects for all comparisons other than CBT vs. PP was statistically significant at −1.255 (−2.713 to −0.470). (Table 4).

Discussion

We identified a fairly dense, well-connected, homogeneous and consistent evidence network around CBT and its control conditions in the acute phase

Table 4. Results of the network meta-regressions adjusting for small study effects

	PP	NT	WL
CBT	1.34 (0.70–2.24)	1.63 (0.78–3.15)	1.79 (0.97–2.87)
PP	–	1.31 (0.54–2.79)	1.44 (0.64–2.78)
NT	–	–	1.22 (0.48–2.42)

OR shows the effectiveness of the intervention on the left over that on the top. 95% credible intervals in parentheses. Resdev = 96.97 (Data points: 98), DIC = 162.2.
PP, psychological *placebo*; NT, no treatment; WL, waiting list; CBT, cognitive-behaviour therapies.

treatment of depression. The effect size estimates for CBT were substantively different, depending on the control condition; the odds ratios for response (50% or greater reduction in depression severity) was not statistically significant at 1.7 (0.8–3.1) when CBT was compared against PP, but were 2.4 (1.3–4.3) in comparison with NT and 6.3 (3.9–10.1) in comparison with WL. Surprisingly, the network meta-analytical estimate of the odds ratio of NT over WL was significantly greater than unity at 2.9 (1.3–5.7).

There are, however, several factors that may undermine the robustness of these estimates. First of all, the methodological standards of the included studies were often less than ideal, as has already been often pointed out by several authors with regard to psychotherapy literature (41, 42). Apart from therapist and participant blinding, which essentially is impossible in psychotherapy trials, only a quarter of the included studies, on average, were rated to be at low risk of bias on various aspects of methodological rigour. We were unable to run the preplanned sensitivity analyses, limiting the included studies to those with high quality. We must remember, however, that these are still all randomized trials satisfying the minimum level of evidence quality and that this is the best evidence body we currently possess.

The apparent existence of small study effects is another major threat. There were notable small study effects in the comparison between CBT and WL. Why small study effects were so preponderant in the comparison against WL is not clear. To gauge the influence of the small study effects bias on our comparison of interest between NT and WL, we ran sensitivity analyses, trying to statistically adjust for such biases through meta-regression and found the odds ratio of NT over WL was no longer statistically significant in the better fitting of the two hypothesized models. However, there is no established method to adjust for funnel plot asymmetry in pairwise let alone network meta-analyses (30). It is possible that our methods may have over-corrected for the small study effects as none of the ORs for CBT over control conditions, including NT and WL, was significant in this adjusted model. These results therefore remain exploratory.

The differential drop-out rates cannot be the reason for the observed difference in effect sizes because we assumed all drop-outs to be non-responders, making the comparison most favourable to WL with the smallest drop-out rate and least favourable to NT with the highest drop-out rate (43). The differences in the drop-out rates would have worked, if any, in decreasing or even subverting the observed differences between NT and WL.

One may wonder how WL can be 'less effective' than NT. In both conditions, participants had earlier shown interest in receiving CBT for their depression, had provided informed consent to be randomized but then allocated to inactive control conditions. Ethically, participants on both NT and WL are allowed to receive some medical care of their own choice during the course of the study on a naturalistic basis. Those allocated to WL may, however, be more motivated to remain depressive so that they can receive their originally desired therapy after the study period is over, while those allocated to NT may more actively seek other treatments, either by oneself or by others, for their ailment.

In summary, the currently available best evidence, analysed by use of NMA, suggested that different control conditions lead to substantively different treatment effect estimates and that WL control may generate bigger effect sizes estimates for CBT than NT or PP. In other words, WL could be regarded a nocebo condition if it indeed is inferior to NT, that is, doing nothing. Unfortunately, the less than ideal quality of the evidence body, including probable publication bias, undermines the strength of this finding. However, there are several research implications of this study that remain valid. We will need to pay more attention to the differences in the control conditions in future psychotherapy research. In individual trials of psychotherapy, the use of WL as control should be more carefully deliberated, as it probably cannot be equated with NT condition controlling for regression towards the mean and the natural course of the disease but instead it may introduce negative psychological expectation of 'waiting for the desired active treatment.' In systematic reviews and meta-analyses of psychotherapies, we probably should not lump different control conditions into one comparative arm.

Acknowledgements

This review is one publication of the High Impact Reviews of Effectiveness in Depression (HIRED) project in which a group of researchers within the Cochrane Collaboration Depression, Anxiety and Neurosis Group conducted systematic reviews of all available evidence for all psychological therapies for treating depression.

Declaration of interest

T. A. Furukawa has received honoraria for speaking at CME meetings sponsored by Asahi Kasei, Eli Lilly, GlaxoSmith-Kline, Mochida, MSD, Otsuka, Pfizer, Shionogi and

Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha. He is on advisory board for Sekisui Chemicals and Takeda Science Foundation. The Japanese Ministry of Education, Science, and Technology, the Japanese Ministry of Health, Labor and Welfare, and the Japan Foundation for Neuroscience and Mental Health have funded his research projects. DC has received honoraria for delivering training at Pfizer. DC is funded by the Medical Research Council, UK. All the other authors have no conflicts of interest to declare.

References

1. Mohr DC, Spring B, Freedland KE et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;**78**:275–284.
2. Borkovec TD, Sibrava NJ. Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. J Clin Psychol 2005;**61**:805–818.
3. Rosenthal R, Frank JD. Psychotherapy and the placebo effect. Psychol Bull 1956;**53**:294–302.
4. 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**:173–178.
5. Jakobsen JC, Hansen JL, Storebo OJ, Simonsen E, Gluud C. The effects of cognitive therapy versus 'no intervention' for major depressive disorder. PLoS ONE 2011;**6**:e28299.
6. Stevens SE, Hynan MT, Allen M. A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. Clin Psychol Sci Pract 2000;**7**:273–290.
7. Hrobjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2010: CD003974.
8. Baker AL, Hiles SA, Thornton LK, Hides L, Lubman DI. A systematic review of psychological interventions for excessive alcohol consumption among people with psychotic disorders. Acta Psychiatr Scand 2012;**126**:243–255.
9. Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. J Consult Clin Psychol 2003;**71**:973–979.
10. Cuijpers P, van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatments of depression: a metaregression analysis. Psychother Res 2008;**18**:225–236.
11. Watanabe N, Hunot V, Omori IM, Churchill R, Furukawa TA. Psychotherapy for depression among children and adolescents: a systematic review. Acta Psychiatr Scand 2007;**116**:84–95.
12. Vancampfort D, Vansteelandt K, Scheewe T et al. Yoga in schizophrenia: a systematic review of randomised controlled trials. Acta Psychiatr Scand 2012;**126**:12–20.
13. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;**15**:2733–2749.
14. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;**159**:130–137.
15. Pinquart M, Duberstein PR, Lyness JM. Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. Am J Psychiatry 2006;**163**:1493–1501.
16. 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**:909–922.
17. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;**23**:56–62.
18. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;**4**:561–571.
19. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;**59**:12–19.
20. 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.
21. da Costa BR, Rutjes AW, Johnston BC et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. Int J Epidemiol 2012;**41**:1445–1459.
22. Shinohara K, Honyashiki M, Imai H et al. Behavioural therapies versus other psychological therapies for depression. Cochrane Database Syst Rev 2013;**10**:CD008696.
23. Higgins JP, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.1.1 [updated March 2011] Available from: www.cochrane-handbook.org, 2011.
24. Dersimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;**7**:177–188.
25. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;**315**:629–634.
26. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;**36**:1–48.
27. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;**50**:683–691.
28. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Series B 2002;**64**:583–639.
29. Dias S, Sutton AJ, Welton NJ, Ades AE. Nice, DSU technical support document 3: heterogeneity: subgroup, meta-regression, bias and bias-adjustment. Available at: http://www.nicedsu.org.uk/TSD3%20Heterogeneity.final%20report.08.05.12.pdf (accessed 26 March 2014).
30. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Methods 2012;**3**:161–176.
31. Thompson LW, Gallagher D, Breckenridge JS. Comparative effectiveness of psychotherapies for depressed elders. J Consult Clin Psychol 1987;**55**:385–390.
32. Comas-Diaz L. Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. J Consult Clin Psychol 1981;**49**: 627–632.
33. Cullen JM. Testing the effectiveness of behavioral activatin therapy in the treatment of acute unipolar depression. [PhD dissertation]. Western Michigan University, 2002.
34. Fry PS. Cognitive training and cognitive-behavioral variables in the treatment of depression in the elderly. Clin Gerontol 1984;**3**:25–45.
35. Morris NE. A group self-instruction method for the treatment of depressed outpatients. [PhD dissertation]. University of Toronto, 1975.
36. Neimeyer RA, Weiss ME. Cognitive and symptomatic predictors of outcome of group therapies for depression. J Cogn Psychother 1990;**4**:23–32.

37. Shaw BF. Comparison of cognitive theerapy and behavior therapy in the treatment of depression. J Consult Clin Psychol 1977;**45**:543–551.
38. Zeiss AM, Lewinsohn PM, Munoz RF. Nonspecific improvement effects in depression using interpersonal skills training, pleasant activity schedules, or cognitive training. J Consult Clin Psychol 1979;**47**:427–439.
39. Propst LR. The comparative efficacy of religious and nonreligious imagery for the treatment of mild depression in religious individuals. Cogn Ther Res 1980;**4**: 167–178.
40. Fuchs CZ, Rehm LP. A self-control behavior therapy program for depression. J Consult Clin Psychol 1977;**45**:206–215.
41. 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**:211–223.
42. 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.
43. Mazzotti E, Barbaranelli C. Dropping out of psychiatric treatment: a methodological contribution. Acta Psychiatr Scand 2012;**126**:426–433.
44. Arean PA, Perri MG, Nezu AM, Schein RL, Christopher F, Joseph TX. Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. J Consult Clin Psychol 1993;**61**:1003–1010.
45. Ayen I, Hautzinger M. Kognitive verhaltenstherapie bei depressionen im klimakterium: eine kontrollierte, randomisierte interventionsstudie. Z Klin Psychol Psychother 2004;**33**:290–299.
46. Barrera M Jr. An evaluation of a brief group therapy for depression. J Consult Clin Psychol 1979;**47**:413–415.
47. Besyner JK. The comparative efficacy of cognitive and behavioral treatments of depression: A multi-assessment approach [PhD dissertation]. Graduate Faculty of Texas Tech University, 1978.
48. Brown RA, Lewinsohn PM. A psychoeducational approach to the treatment of depression: comparison of group, individual, and minimal contact procedures. J Consult Clin Psychol 1984;**52**:774–783.
49. Carrington CH. A comparison of cognitive and analytically oriented brief treatment approaches to depression in black women [PhD dissertation]. University of Maryland, 1979.
50. Collins RW. The treatment of depression: An integrative psychotherapy model [PhD dissertation]. San Francisco, CA: Saybrook Institute, 1996.
51. Dowrick C, Dunn G, Ayuso-Mateos JL et al. Problem solving treatment and group psychoeducation for depression: multicentre randomised controlled trial. Outcomes of Depression International Network (ODIN) Group. BMJ 2000;**321**:1450–1454.
52. Embling S. The effectiveness of cognitive behavioural therapy in depression. Nurs Stand 2002;**17**:33–41.
53. Epstein D. Aerobic activity versus group cognitive therapy: An evaluative study of contrasting interventions for the alleviation of clinical depression [PhD dissertation]. University of Nevada Reno, 1986.
54. Faramarzi M, Alipor A, Esmaelzadeh S, Kheirkhah F, Poladi K, Pash H. Treatment of depression and anxiety in infertile women: cognitive behavioral therapy versus fluoxetine. J Affect Disord 2008;**108**:159–164.
55. Hamamci Z. Integrating psychodrama and cognitive behavioral therapy to treat moderate depression. Arts Psychother 2006;**33**:199–207.
56. Hamdan-Mansour AM, Puskar K, Bandak AG. Effectiveness of cognitive-behavioral therapy on depressive symptomatology, stress and coping strategies among Jordanian university students. Issues Ment Health Nurs 2009;**30**: 188–196.
57. Hautzinger M, Welz S. Kognitive verhaltenstherapie bei depressionen im alter. Ergebnisse einer kontrollierten vergleichsstudie unter ambulanten bedingungen an depressionen mittleren schweregrads [Cognitive behavioral therapy for depressed older outpatients–a controlled, randomized trial]. Z Gerontol Geriatr 2004;**37**:427–435.
58. Hayman PM, Cope CS. Effects of assertion training on depression. J Clin Psychol 1980;**36**:534–543.
59. Hegerl U, Hautzinger M, Mergl R 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.
60. Hess-Homeier MJ. A comparison of Beck's cognitive therapy and jogging as treatments for depression [PhD dissertation]. University of Montana, 1981.
61. Kelly LM. Rational emotive therapy versus Lewinsohnian based approaches to the treatment of depression [PhD dissertation]. Athens, GA: University of Georgia, 1982.
62. Malouff JM. A study of a brief, cognitive treatment for depression personsl who have recently experienced a martial separation [PhD dissertation]. Arizona State University, 1984.
63. Miranda J, Chung JY, Green BL et al. Treating depression in predominantly low-income young minority women: a randomized controlled trial. JAMA 2003;**290**:57–65.
64. Nezu AM. Efficacy of a social problem-solving therapy approach for unipolar depression. J Consult Clin Psychol 1986;**54**:196–202.
65. Nezu AM, Perri MG. Social problem-solving therapy for unipolar depression: an initial dismantling investigation. J Consult Clin Psychol 1989;**57**:408–413.
66. Pace FR. Behavioral techniques in the treatment of depression [PhD dissertation]. University of New South Wales, 1977.
67. Pace TM. Changes in depressive self-schemata and depressive symptoms following cognitive therapy. J Couns Psychol 1993;**40**:288–294.
68. Pecheur DR. A comparison of the efficacy of secular and religious cognitive behavior modification in the treatment of depressed Christian college students [PhD dissertation]. Rosemead Graduate School of Professional Psychology, 1980.
69. Pellowe ME. Acceptance and commitment therapy as a treatment for dysphoria [PhD dissertation]. University of Wyoming, 2006.
70. Propst LR, Ostrom R, Watkins P, Dean T, Mashburn D. Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. J Consult Clin Psychol 1992;**60**:94–103.
71. Ross M, Scott M. An evaluation of the effectiveness of individual and group cognitive therapy in the treatment of depressed patients in an inner city health centre. JR Coll Gen Pract 1985;**35**:239–242.
72. Schmidt MM, Miller WR. Amount of therapist contact and outcome in a multidimensional depression treatment program. Acta Psychiatr Scand 1983;**67**:319–332.