Critical appraisal of randomised controlled trials

Dr Kamal R. Mahtani
BSc PhD MBBS PGDip MRCGP
GP and Clinical Lecturer
Centre for Evidence Based Medicine
University of Oxford

November 2014
Sickness in Salonica: my first, worst, and most successful clinical trial-1941.

“. . . I recruited 20 young prisoners . . . I gave them a short talk about my medical hero James Lind and they agreed to co-operate in an experiment. I cleared two wards. I numbered the 20 prisoners off: odd numbers to one ward and evens to the other.

Each man in one ward received two spoonfuls of yeast daily. The others got one tablet of vitamin C from my "iron" reserve. The orderlies co-operated magnificently . . . They controlled fluid intake and measured frequency of urination.

. . . There was no difference between the wards for the first two days, but the third day was hopeful, and on the fourth the difference was conclusive . . . there was less oedema in the "yeast" ward. I made careful notes of the trial and immediately asked to see the Germans.”

“It could be argued that the trial was randomised and controlled, although this last was somewhat inadequate. In those early days, when the randomised controlled trial was little known in medicine, this was something of an achievement.”
STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS
A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor C. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D'Arcy Hart (secretary). The centres at which the work was carried out and the specialists in charge of patients and pathological work were as follows:

**Brompton Hospital, London.**—Clinician: Dr. J. W. Crofton, Streptomycin Registrar (working under the direction of the honorary staff of Brompton Hospital); Pathologists: Dr. J. W. Clegg, Dr. D. A. Mitchison.

**Colindale Hospital (L.C.C.), London.**—Clinicians: Dr. J. V. Hurford, Dr. B. J. Douglas Smith, Dr. W. E. Snell; Pathologists (Central Public Health Laboratory): Dr. G. B. Forbes, Dr. H. D. Holt.

**Harefield Hospital (M.C.C.), Harefield, Middlesex.**—Clinicians: Dr. R. H. Brent, Dr. L. E. Houghton; Pathologist: Dr. E. Nassau.

**Bangour Hospital, Bangour, West Lothian.**—Clinician: Dr. I. D. Ross; Pathologist: Dr. Isabella Purdie.

**Killingbeck Hospital and Sanatorium, Leeds.**—Clinicians: Dr. W. Santon Gilmour, Dr. A. M. Reevie; Pathologist: Professor J. W. McLeod.

**Northern Hospital (L.C.C.), Winchmore Hill, London.**—Clinicians: Dr. F. A. Nash, Dr. R. Shoulman; Pathologists: Dr. J. M. Alston, Dr. A. Mohun.

**Sully Hospital, Sully, Glam.**—Clinicians: Dr. D. M. E. Thomas, Dr. L. R. West; Pathologist: Professor W. H. Tytler.

The clinicians of the centres met periodically as a working subcommittee under the chairmanship of Dr. Geoffrey Marshall; so also did the pathologists under the chairmanship of Dr. R. Cruickshank. Dr. Marc Daniels, of the Council’s scientific staff, was responsible for the clinical co-ordination of the trials, and he also prepared the report for the Committee, with assistance from Dr. D. A. Mitchison on the analysis of laboratory results. For the purpose of final analysis the radiological findings were assessed by a panel composed of Dr. L. G. Blair, Dr. Peter Kerley, and Dr. Geoffrey S. Todd.

**Introduction**

When a special committee of the Medical Research Council undertook in September, 1946, to plan clinical trials of streptomycin in tuberculosis the main problem faced was that of investigating the effect of the drug in pulmonary tuberculosis. This antibiotic had been discovered two years previously by Waksman (Schatz, Bugie, and Waksman, 1944); in the intervening period its power of inhibiting tubercle bacilli was well established. but if based on adequately controlled clinical trials (Hinshaw and Feldman, 1944). The one controlled trial of gold treatment (and the only report of an adequately controlled trial in tuberculosis we have been able to find in the literature) reported negative therapeutic results (Amberston, McMahon, and Pinner, 1931). In 1946 no controlled trial of streptomycin in pulmonary tuberculosis had been undertaken in the U.S.A. The Committee of the Medical Research Council decided then that a part of the small supply of streptomycin...
What's so special about RCTs?

• most rigorous way of determining:
  – a cause-effect relation exists between treatment and outcome and
  – for assessing the cost effectiveness of a treatment

• distributing the characteristics of patients that may influence the outcome randomly between the groups-no systematic differences between intervention groups
What's so special about RCTs?

- patients and trialists should remain unaware of which treatment was given until the study is completed to avoid influencing the result
- both arms treated identically except for the intervention of interest – estimating the size of the difference in predefined outcomes between intervention groups
So are RCTs the gold standard for evidence?

…..depends
Limitations of RCTs

• Excellent vs Poor RCTs – quality varies
  – Impact on interpretation of result (external validity)?

• Expensive and time consuming
  – £250k - £millions over 2-5 years+

• May not always be the right study design to answer that question
Practicing EBM – the 4 A’s

Step 1: Ask a clinical question
Step 2: Acquire the best evidence
Step 3: Appraise the evidence
Step 4: Apply the evidence
Levels of evidence

- Systematic Reviews
- Randomized Controlled Trials
- Cohort Studies
- Case-Control Studies
- Case Series, Case Reports
- Editorials, Expert Opinion
Practicing EBM – the 4 A’s

Step 1: Ask a clinical question
Step 2: Acquire the best evidence
Step 3: Appraise the evidence
Step 4: Apply the evidence
Types of evidence

Critical appraisal
Risk of Bias

The degree to which the result is skewed away from the truth
Internal validity

• extent to which observed treatment effects can be ascribed to differences in treatment and not confounding, thereby allowing the inference of causality to be ascribed to a treatment.¹

• Systematic error (bias) could threaten the internal validity of trials, and all efforts should be made to minimise these in the design, conduct, and analysis of studies.²

1. http://www.bmj.com/content/344/bmj.e1004
Confounding factors

- Other patient features/causal factors, apart from the one being measured, that can affect the outcome of the study e.g..
External validity

• The degree to which the results of the study can be applied to other populations
Assessing risk of bias for an RCT

Critical Appraisal for Therapy Articles

**THERAPY STUDY:** Are the results of the trial valid? (Internal Validity)
What question did the study ask?

<table>
<thead>
<tr>
<th>Patients -</th>
<th>Intervention -</th>
<th>Comparison -</th>
<th>Outcome(s) -</th>
</tr>
</thead>
</table>

**1a. R- Was the assignment of patients to treatments randomised?**

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centralised computer randomisation is ideal and often used in multi-centred trials. Smaller trials may use an independent person (e.g., the hospital pharmacy) to “police” the randomization.</td>
<td>The Methods should tell you how patients were allocated to groups and whether or not randomisation was concealed.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐
Comment: |

**1b. R- Were the groups similar at the start of the trial?**

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>If the randomisation process worked (that is, achieved comparable groups) the groups should be similar. The more similar the groups the better it is. There should be some indication of whether differences between groups are statistically significant (i.e., p values).</td>
<td>The Results should have a table of “Baseline Characteristics” comparing the randomized groups on a number of variables that could affect the outcome (e.g., age, risk factors etc). If not, there may be a description of group similarity in the first paragraphs of the Results section.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐
Comment: |

**2a. A - Aside from the allocated treatment, were groups treated equally?**

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apart from the intervention the patients in the different groups should be treated the same, e.g., additional treatments or tests.</td>
<td>Look in the Methods section for the follow-up schedule, and permitted additional treatments, etc and in Results for actual use.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐
Comment: |

**2b. A - Were all patients who entered the trial accounted for? - and were they analysed in the groups to which they were randomised?**

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losses to follow-up should be minimal — preferably less than 20%. However if few patients have the outcome of interest then randomization is likely to be important.</td>
<td>The Results section should say how many patients were randomised (e.g., Baseline Characteristics table) and how...</td>
</tr>
</tbody>
</table>
Depression Management

Risk and f/u

Pharmacological
- TCA
- SSRI
- SNRI

Non-pharmacological
- Psychological therapies
  - Mindfulness group
  - Psychodynamic therapy
  - Behavioural activation
  - Individual CBT
- Self help and lifestyle modification
  - Alcohol, diet, social networks, sleep
- Structured exercise
taking regular physical exercise
Exercise does little to help the symptoms of depression, new study finds

By SUZANNAH HILLS

PUBLISHED: 08:23, 6 June 2012 | UPDATED: 11:01, 6 June 2012

Exercise does little to help alleviate the symptoms of depression, a new study has found.

The findings contrast with current clinical guidance which recommends exercise to help those suffering from the mental illness that affects one in six adults in Britain at any one time.

But research published in the British Medical Journal suggests that doing a physical activity combined with usual treatment did not reduce symptoms of depression more than the treatment alone.
Facilitated physical activity as a treatment for depressed adults: randomised controlled trial

Melanie Chalder research fellow1, Nicola J Wiles senior lecturer1, John Campbell professor2, Sandra P Hollinghurst senior lecturer3, Anne M Haase senior lecturer4, Adrian H Taylor professor2, Kenneth R Fox professor2, Ceire Costelloe research associate1, Aidan Searle research associate1, Helen Baxter research associate5, Rachel Winder associate research fellow1, Christine Wright associate research fellow1, Katrina M Turner lecturer1, Michael Caiman professor3, Deborah A Lawlor professor4, Tim J Peters professor5, Deborah J Sharp professor5, Alan A Montgomery reader1, Glyn Lewis professor1

1School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK; 2Primary Care Research Group, Peninsula Medical School, Exeter, UK; 3School of Policy Studies, University of Bristol; 4Sport and Health Sciences, University of Exeter; 5School of Social Policy, University of Kent, Canterbury, UK; 6School of Clinical Sciences, University of Bristol

Abstract

Objective To investigate the effectiveness of facilitated physical activity as an adjuvant treatment for adults with depression presenting in primary care.

Design Pragmatic, multicentre, two-arm parallel randomised controlled trial

Setting General practices in Bristol and Exeter.

Participants 361 adults aged 18-69 who had recently consulted their general practitioner with symptoms of depression. All those randomised had a diagnosis of an episode of depression as assessed by the clinical interview schedule-revised and a Beck depression inventory score of 14 or more.

Interventions In addition to usual care, intervention participants were offered up to three face-to-face sessions and 10 telephone calls with a trained physical activity facilitator over eight months. The intervention was based on theory and aimed to provide individually tailored support and encouragement to engage in physical activity.

Main outcome measures The primary outcome was self-reported symptoms of depression, assessed with the Beck depression inventory at four months post-randomisation. Secondary outcomes included use of antidepressants and health-related quality of life at the four-month time-point.

Results At 12 months, the intervention group reported a decrease in the Beck depression inventory score of 14.54 (95% confidence interval 12.19 to 16.89; P=0.005) compared with the usual care group. Similarly, there was no evidence that the intervention group reported a change in mood by the eighth or 12 month follow-up points. Nor was there evidence that the intervention reduced antidepressant use compared with usual care (adjusted odds ratio 0.63, 95% confidence interval 0.49 to 0.96; P=0.03) over the duration of the trial. However, participants allocated to the intervention group reported a decrease in physical activity during the follow-up period than those allocated to the usual care group (adjusted odds ratio 2.27, 95% confidence interval 1.32 to 3.89; P<0.003).

Conclusions The addition of a facilitated physical activity intervention to usual care did not improve depression outcome or reduce use of antidepressants compared with usual care alone.

Trial registration Current Controlled Trials ISRCTN16900744

Introduction

Depression is one of the most common reasons for consulting a general practitioner within the United Kingdom, and its associated economic burden is considerable. Although antidepressants are effective, many patients and healthcare professionals would like other options to be available as an alternative or adjunct to drug therapy. Some evidence shows
Critical appraisal....

...is like being a detective.

You need the skills to think broadly and detect the flaws that might distract you from finding the true answer.
Recruitment (selection bias)

General population

- Sample population
- Target population
- Sample population
Recruitment (selection bias)

• Were the subjects representative of the target population?
  – What were the inclusion & exclusion criteria?
  – Were they appropriate?
  – How/where were they recruited from?

• Methods Recruitment of participants and baseline assessment & Results 1st para
Randomisation (selection bias)
Allocation concealment

How was the randomised sequence implemented?

BEST – most valid technique
- Central computer randomization

DOUBTFUL
- Envelopes, etc
Allocation (selection bias)

- Were the groups comparable at the start?
  - “Table 1”
- Randomised appropriately?
- Allocation to group concealed beforehand?

- Methods: Randomisation, concealment, and blinding and “Table 1”
Maintenance

• Were both groups comparable throughout the study?
  – Managed equally bar the intervention?
    • What was the intervention?
    • What was the comparator?

• Methods: Follow up and Intervention and comparator (usual care)
Adequate follow up? (Attrition bias)
Adequate follow up? (Attrition bias)

- How many people were lost to f/u?
- Why were they lost to f/u?
- Did the researchers use an intention to treat (ITT) principle?
  - Once a participant is randomised, they should be analysed to the group they were assigned to
- Figure 1 and Statistical analysis
Measurement – blinding (Performance bias)

**SINGLE BLIND**
- Patients: aware of the test being conducted.
- Doctor: unaware of the test being conducted.
- Control Pills: regular.
- Test Pills: placebo.

**DOUBLE BLIND**
- Patients: unaware of the test being conducted.
- Doctor: aware of the test being conducted.
- Control Pills: regular.
- Test Pills: placebo.

**UNBLINDED**

Measurement – blinding (Performance bias)

• Were the outcomes measured blindly by researchers and participants?

• Methods: Randomisation, concealment, and blinding
Statistics
**P - values and CI**

- **P values**
  - Measure of probability that a result is due to chance
  - The smaller the value (usually P<0.05) less likely due to chance

- **Confidence intervals**
  - Estimate of the range of values that are likely to include the real value
  - 95% chance of including the real value
  - Narrower the range>more reliable
  - If value does not cross 0 for a difference, or 1 for a ratio then pretty sure result is real (p<0.05)
Measurement - outcomes

• What were the outcomes?
  – Primary
  – Secondary
  – Were they appropriate?

• How were the results reported?

• Were they significant?

• Methods: Outcomes and Results
1. Sadness
   0. I do not feel sad.
   1. I feel sad much of the time.
   2. I am sad all the time.
   3. I am so sad or unhappy that I can't stand it.

2. Pessimism
   0. I am not discouraged about my future.
   1. I feel more discouraged about my future than I used to.
   2. I do not expect things to work out for me.
   3. I feel my future is hopeless and will only get worse.

3. Past Failure
   0. I do not feel like a failure.
   1. I have failed more than I should have.
   2. As I look back, I see a lot of failures.
   3. I feel I am a total failure as a person.

4. Loss of Pleasure
   0. I get as much pleasure as I ever did from the things I enjoy.
   1. I don't enjoy things as much as I used to.
   2. I get very little pleasure from the things I used to enjoy.
   3. I can't get any pleasure from the things I used to enjoy.

5. Guilty Feelings
   0. I don't feel particularly guilty.
   1. I feel guilty over many things I have done or should have done.
   2. I feel quite guilty most of the time.
   3. I feel guilty all of the time.

6. Punishment Feelings
   0. I don't feel I am being punished.
   1. I feel I may be punished.
   2. I expect to be punished.
   3. I feel I am being punished.

7. Self-Dislike
   0. I feel the same about myself as ever.
   1. I have low confidence in myself.
   2. I am disappointed in myself.
   3. I dislike myself.

8. Self-Criticalness
   0. I don't criticize or blame myself more than usual.
   1. I am more critical of myself than I used to be.
   2. I criticize myself for all of my faults.
   3. I blame myself for everything bad that happens.

9. Suicidal Thoughts or Wishes
   0. I don't have any thoughts of killing myself.
   1. I have thoughts of killing myself, but I would not carry them out.
   2. I would like to kill myself.
   3. I would kill myself if I had the chance.

10. Crying
    0. I don't cry anymore than I used to.
    1. I cry more than I used to.
    2. I cry over every little thing.
    3. I feel like crying, but I can't.

11. Agitation
    0. I am no more restless or wound up than usual.
    1. I feel more restless or wound up than usual.
    2. I am so restless or agitated that it's hard to stay still.
    3. I am so restless or agitated that I have to keep moving or doing something.

12. Loss of Interest
    0. I have not lost interest in other people or activities.
    1. I am less interested in other people or things than before.
    2. I have lost most of my interest in other people or things.
    3. It's hard to get interested in anything.

13. Indecisiveness
    0. I make decisions about as well as ever.
    1. I find it more difficult to make decisions than usual.
    2. I have much greater difficulty in making decisions than I used to.
    3. I have trouble making any decisions.

14. Worthlessness
    0. I do not feel I am worthless.
    1. I don't consider myself as worthwhile and useful as I used to.
    2. I feel more worthless as compared to other people.
    3. I feel utterly worthless.

15. Loss of Energy
    0. I have as much energy as ever.
    1. I have less energy than I used to.
    2. I don't have enough energy to do very much.
    3. I don't have enough energy to do anything.

16. Changes in Sleeping Pattern
    0. I have not experienced any change in my sleeping pattern.
    1a. I sleep somewhat more than usual.
    1b. I sleep somewhat less than usual.
    2a. I sleep a lot more than usual.
    2b. I sleep a lot less than usual.
    3a. I sleep most of the day.

17. Irritability
    0. I am no more irritable than usual.
    1. I am much more irritable than usual.
    2. I am irritated all the time.

18. Changes in Appetite
    0. I have not experienced any change in my appetite.
    1a. My appetite is somewhat less than usual.
    1b. My appetite is somewhat greater than usual.
    2a. My appetite is much less than before.
    2b. My appetite is much greater than usual.
    3a. I have no appetite at all.
    3b. I crave food all the time.

19. Concentration Difficulty
    0. I can concentrate as well as ever.
    1. I can't concentrate as well as usual.
    2. It's hard to keep my mind on anything for very long.
    3. I find I can't concentrate on anything.

20. Tiredness or Fatigue
    0. I am more tired or fatigued than usual.
    1. I get more tired or fatigued more easily than usual.
    2. I am too tired or fatigued to do a lot of the things I used to do.
    3. I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex
    0. I have not noticed any recent change in my interest in sex.
    1. I am less interested in sex than I used to be.
    2. I am much less interested in sex now.
    3. I have lost interest in sex completely.
# Outcomes

<table>
<thead>
<tr>
<th>Measure</th>
<th>Narrative</th>
<th>Numerical</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary outcome:</strong> short term symptoms of depression</td>
<td>Beck depression inventory score</td>
<td>no evidence that participants in the intervention group had a better outcome at four months than those in the usual care group</td>
</tr>
<tr>
<td><strong>Secondary outcomes</strong>&lt;br&gt;Longer term symptoms of depression</td>
<td>Beck depression inventory score</td>
<td>no evidence of a difference between the treatment groups over the duration of the study</td>
</tr>
<tr>
<td>Anti-depressant use</td>
<td>participants reporting use of antidepressants</td>
<td>no evidence to suggest any difference between the groups at either the four month follow-up point or duration of trial</td>
</tr>
<tr>
<td>Physical activity</td>
<td>self completion seven day recall diary</td>
<td>there was some evidence for a difference in reported physical activity between the groups at four months post-randomisation</td>
</tr>
</tbody>
</table>
Conclusions of the study

What is already known on this topic

Depression is a leading contributor to disability in the United Kingdom and is associated with a decrement of health greater than many other chronic diseases.

Many patients and healthcare professionals would like an effective and accessible non-drug treatment for depression.

Numerous studies have reported the positive effects of physical activity but most of the current evidence originates from small non-clinical samples using interventions that are not practicable in healthcare settings.

What this study adds

A physical activity intervention in addition to usual care did not improve symptoms of depression or reduce the use of antidepressants compared with usual care alone.

The intervention increased self reported physical activity and this effect was sustained for 12 months.

Clinicians and policy makers should alert people with depression that advice to increase physical activity will not increase their chances of recovery from depression.

Cite this as: BMJ 2012;344:e2758

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non-commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-
External validity/applicability

Would you advocate exercise for depression based on this study?
Exercise ‘fails to lift clinical depression’

Exercise should not be “prescribed” to people with clinical depression, according to a study which found it did nothing to improve their moods.
Exercise ‘no help for depression’ research suggests

By Branwen Jeffrey
Health correspondent, BBC News

Combining exercise with conventional treatments for depression does not improve recovery, research suggests.

In the NHS-funded study - published in the British Medical Journal - some patients were given help to boost their activity levels in addition to receiving therapy or antidepressants.

After a year at 361 patients had fewer signs of depression, but there was no difference between the two groups.

Current guidelines suggest sufferers do up to three exercise sessions a week.

The National Institute for Health and Clinical Excellence (Nice) drew up that advice in 2004.

At the time it cast that on the basic of the research available. nice/press
Summary

- Lots of “evidence” in healthcare
- RCTs provide an opportunity to deliver answers to the effects of interventions
- But dependent upon minimizing risk of bias
- Critical appraisal assess this
- Lots of tools to assess risk of bias
- Application (external validity) based on your interpretation of results
Want more?
RCT course

https://www.conted.ox.ac.uk/
Thank you

kamal.mahtani@phc.ox.ac.uk

@krmahani
Group work
Exercise for depression: critical appraisal

- 2-3 groups
- 2-3 different RCTs from same SR
- In groups:
  - Read paper – DON’T REFER BACK TO COCHRANE RV!
  - PICO
  - Critical appraisal – internal validity
  - External validity
  - Each group present their paper (PICO, appraisal)
  - Comment on the validity for 10 mins
For those with mental health problems, exercise can be more effective than antidepressants

- Exercise raises self-esteem and boosts levels of 'feel good' hormones
- For those with mild depression, physical activity improves symptoms

As a Zumba instructor, Karen Bedford might be expected to rave about what she teaches. But Karen has a more personal motive, for she believes exercise helped her recover from a decade-long battle with depression.

"Since Zumba became part of my life my mental state's been more stable and I haven't needed the pills," says Karen, 49, who has two grown-up children.

"I get the odd down day, but because I've regained my confidence I face my problems much better now."

It's a far cry from when Karen's depression emerged in 1996. Her long-term relationship had just ended and she was fighting to keep the family home in Potters Bar, Hertfordshire, when she was made redundant from her job at a publishing house.
<table>
<thead>
<tr>
<th>Method</th>
<th>RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>University students aged 18 - 25 with depression 100% women</td>
</tr>
<tr>
<td>Interventions</td>
<td>1. 40 - 60 minutes of running, 3 times a week, supervised. (n = 10) 2. Control group with no active intervention (n = 10)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Beck Depression Inventory score</td>
</tr>
<tr>
<td>Notes</td>
<td>Small sample size (10 participants in each arm); specific population under study</td>
</tr>
</tbody>
</table>
# Hemat-Far 2012

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>High risk</td>
<td>Clinician judgement used at recruitment. After reviewing questionnaires psychiatrists “selected” 20 women</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk</td>
<td>No information given</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) participants</td>
<td>Unclear risk</td>
<td>Participants not blinded to intervention; unclear effect on bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) those delivering intervention</td>
<td>Unclear risk</td>
<td>No information given</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) outcome assessors</td>
<td>High risk</td>
<td>Self report BDI</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias) All outcomes</td>
<td>Unclear risk</td>
<td>No discussion on attrition rate</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Low risk</td>
<td>BDI specified at outset and completed in results</td>
</tr>
<tr>
<td>Sims 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>RCT</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Recruited from hospital databases of stroke patients discharged in last year, general practitioners and newspaper articles. Had to be over 6 months post-stroke and have depression confirmed by a psychiatrist. Mean age 67.13 (range 21 to 93), 40% women, N = 45</td>
<td></td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td>1. Group-based moderate-intensity strengthening exercises twice a week for 10 weeks. The PRT programme included 2 high-intensity sessions/week for 10 weeks at a community-based gymnasium. (n = 23) 2. Usual care (n = 22)</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Centre for Epidemiologic Studies for Depression scale</td>
<td></td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>Intention-to-treat analysis, Outcome was self-rated symptoms of depression by CES-D scale</td>
<td></td>
</tr>
</tbody>
</table>
## Sims 2009

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors' judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Block randomised list</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>Randomisation was conducted centrally by an independent person</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) participants</td>
<td>Unclear risk</td>
<td>Participants not blind, unclear risk of bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) those delivering intervention</td>
<td>Unclear risk</td>
<td>Those delivering intervention were not blind, unclear risk of bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) outcome assessors</td>
<td>High risk</td>
<td>Self report outcome (depressive symptoms by CES-D scale)</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias) All outcomes</td>
<td>High risk</td>
<td>Baseline assessment was performed in 45 people; complete data were available for 43 people at 6 months (23/23 in intervention group and 20/22 in the control)</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
<td>Reported all prespecified outcome (though we do not have access to the protocol)</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
<td>Unclear</td>
</tr>
</tbody>
</table>
## Singh 2005

<table>
<thead>
<tr>
<th>Methods</th>
<th>RCT</th>
</tr>
</thead>
</table>
| Participants| People responding to a postal questionnaire who had DSM-IV depression or dysthymia  
Mean age 69  
55% women  
N = 60 |
| Interventions| 1. Progressive resistance training at 80% of 1 repetition max (n = 20)  
2. Resistance training at 20% of 1 repetition max (n = 20)  
3. Usual care (n = 20)  
Each intervention group held 3 times a week for 8 weeks |
| Outcomes    | 1. Hamilton Rating Scale for depression  
2. Geriatric Depression score |
| Notes       | Not intention-to-treat (50/60 completed the study and were available for assessment)  
Outcome assessment blind |
## Singh 2005

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th>Authors’ judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Computer-generated random numbers</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>Adequate. Sealed opaque envelopes open after baseline assessment</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) participants</td>
<td>Unclear risk</td>
<td>Participants not blind, unclear effect on bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) those delivering intervention</td>
<td>Unclear risk</td>
<td>Those delivering the intervention were not blind to treatment allocation, unclear effect on bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) outcome assessors</td>
<td>Low risk</td>
<td>HRSD performed by blinded outcome assessors</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias) All outcomes</td>
<td>High risk</td>
<td>6/60 dropped out (2 from the high-dose, 3 from the low-dose and 1 from the usual care group)</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
<td>Prespecified outcomes in paper were reported, but no protocol</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
<td>Unclear</td>
</tr>
</tbody>
</table>
## Krogh 2009

<table>
<thead>
<tr>
<th>Methods</th>
<th>RCT (parallel group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>Referred from general practitioners, private psychiatrists, psychologists and psychiatric wards institutions. Included if met criteria for major depression</td>
</tr>
<tr>
<td></td>
<td>Mean age 38.9</td>
</tr>
<tr>
<td></td>
<td>73.9% women</td>
</tr>
<tr>
<td></td>
<td>N = 165</td>
</tr>
<tr>
<td>Interventions</td>
<td>1. Strength circuit training (n = 55)</td>
</tr>
<tr>
<td></td>
<td>2. Aerobic (machine-based) training (n = 55)</td>
</tr>
<tr>
<td></td>
<td>3. Relaxation control (n = 55)</td>
</tr>
<tr>
<td></td>
<td>Twice-weekly intervention for 32 sessions delivered over a 4-month period</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Hamilton Rating Scale for Depression</td>
</tr>
<tr>
<td>Notes</td>
<td>Intention-to-treat analysis</td>
</tr>
<tr>
<td></td>
<td>Significant drop-outs in each group</td>
</tr>
<tr>
<td></td>
<td>Changed sample size calculation after first 50 participants on basis of observed standard deviation</td>
</tr>
</tbody>
</table>
## Krogh 2009

<table>
<thead>
<tr>
<th>Bias</th>
<th>Authors' judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk</td>
<td>Computerised restricted randomisation with a block size of 8</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk</td>
<td>The block size and allocation sequence were unknown to the DEMO trial staff</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) participants</td>
<td>Unclear risk</td>
<td>Participants not blind, but unclear what influence this had on bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) those delivering intervention</td>
<td>Unclear risk</td>
<td>Physiotherapists delivering the intervention were not blind. Unclear how this influenced risk of bias</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias) outcome assessors</td>
<td>Low risk</td>
<td>The assessor was blind to intervention group. The investigators asked the outcome assessors to guess intervention group. The kappa values for agreement between the right allocation and the guessed allocation were 0.15 and 0.05 for the assessments at 4 and 12 months respectively</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias) All outcomes</td>
<td>Low risk</td>
<td>137/165 were available for follow-up at the end of the intervention. Eighteen were lost to follow-up and 10 refused to participate (8/55 in strength group, 7/55 in aerobic group and 13/55 in the relaxation group). The authors used a likelihood-based mixed-effect model with an unstructured variance matrix available in SPSS, which is able to handle missing data with higher precision and power than last observation carried forward. The authors reported no significant difference between missing participants and participants included in the analyses at either 4 or 12 months, and concluded that it was reasonable to assume that the missing data were 'missing at random'</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Low risk</td>
<td>All prespecified outcomes seem to have been reported. Protocol was published in advance of the trial</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
<td>The authors repeated power calculations part-way through the trial, and reduced the sample size as the standard deviation was lower than anticipated</td>
</tr>
</tbody>
</table>
### Chu 2008

<table>
<thead>
<tr>
<th>Methods</th>
<th>RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>Volunteers aged 18 to 45 recruited via flyers and word of mouth from University and local physician referral. Depression severity mild to moderate, if severe required written permission from physician. Mean age 26.4 (18 to 43). 100% women.</td>
</tr>
<tr>
<td>Interventions</td>
<td>For 10 weeks: 1. Up to 5 high-intensity aerobic exercise sessions per week (1 supervised) to expend 1000 Kcal per week (n = 15) 2. Up to 5 low-intensity aerobic exercise sessions per week (1 supervised) to expend 1000 Kcal per week (n = 11) 3. Met with investigator once per week for 30 minutes of group stretching exercises (n = 12)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Beck Depression Inventory-II</td>
</tr>
<tr>
<td>Notes</td>
<td>Analysis not intention-to-treat BDI-II self-rated depression score</td>
</tr>
<tr>
<td>Bias</td>
<td>Authors’ judgement</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>participants</td>
<td></td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>those delivering intervention</td>
<td></td>
</tr>
<tr>
<td>Blinding (performance bias and detection bias)</td>
<td>High risk</td>
</tr>
<tr>
<td>outcome assessors</td>
<td></td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>High risk</td>
</tr>
<tr>
<td>All outcomes</td>
<td></td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Other bias</td>
<td>Unclear risk</td>
</tr>
</tbody>
</table>
KEEP CALM IT'S LUNCH TIME
Odds ratio

• odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure

• Interpreting OR
  – OR=1 Exposure does not affect odds of outcome
  – OR>1 Exposure associated with higher odds of outcome
  – OR<1 Exposure associated with lower odds of outcome

• E.g.... OR = 1.46
  – Odds of having the outcome are 1.46 higher in the exposed group vs control group
Odds ratio

\[ \text{OR} = \frac{a/c}{b/d} \]

<table>
<thead>
<tr>
<th>Exposure of interest</th>
<th>Outcome of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>d</td>
</tr>
</tbody>
</table>
Relative Risk or Risk Ratio

- the risk of the event in one group divided by the risk of the event in the other group

- Interpreting RR
  - RR = 1 Exposure does not affect risk of outcome
    - Is the treatment intended to prevent an undesirable outcome?
      - RR < 1 Exposure reduces the risk of the event
      - RR > 1 Exposure increases the risk of the event (possible treatment harm, adverse events)
    - Is the treatment intended to promote an outcome? (e.g. disease remission)
      - RR < 1 Exposure reduces the risk of the event (disease remission)
      - RR > 1 Exposure increases the risk of the event (disease remission)

E.g.... RR = 0.46
- Risk of getting the outcome with the exposure was 0.46 of that in the control group
RR v OR

- Often similar when event rate is low (<10%) or treatment effect is small (close to 1)
- As event rate increases (>10%)
Relative Risk or Risk Ratio

\[ RR = \frac{a/(a+b)}{c/(c+d)} \]
Odds ratio

- odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure

- Interpreting OR
  - OR=1 Exposure does not affect odds of outcome
  - OR>1 Exposure associated with higher odds of outcome
  - OR<1 Exposure associated with lower odds of outcome

- E.g.... OR = 1.46
  - Odds of having the outcome are 1.46 higher in the exposed group vs control group
### Odds ratio

\[
\text{OR} = \frac{a/c}{b/d}
\]

<table>
<thead>
<tr>
<th>Exposure of interest</th>
<th>Outcome of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
</tr>
<tr>
<td>-</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>d</td>
</tr>
</tbody>
</table>
Relative Risk or Risk Ratio

• the risk of the event in one group divided by the risk of the event in the other group
• Interpreting RR
  • RR = 1 Exposure does not affect risk of outcome
    – Is the treatment intended to prevent an undesirable outcome?
      • RR < 1 Exposure reduces the risk of the event
      • RR > 1 Exposure increases the risk of the event (possible treatment harm, adverse events)
    – Is the treatment intended to promote an outcome? (e.g. disease remission)
      • RR < 1 Exposure reduces the risk of the event (disease remission)
      • RR > 1 Exposure increases the risk of the event (disease remission)

E.g.…. RR = 0.46

  – Risk of getting the outcome with the exposure was 0.46 of that in the control group
RR v OR

• Often similar when event rate is low (<10%) or treatment effect is small (close to 1)
• As event rate increases (>10%)
Relative Risk or Risk Ratio

<table>
<thead>
<tr>
<th></th>
<th>Outcome of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure of interest</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
</tr>
</tbody>
</table>

$$RR = \frac{a/(a+b)}{c/(c+d)}$$
Selection bias

• systematic differences between baseline characteristics of the groups

• Adequate randomisation
  – 1) Sequence generation
  – 2) Allocation concealment
Sequence generation (selection bias)

Low risk of bias
- random number table
- Using a computer random number generator
- Coin tossing
- Shuffling cards or envelopes
- Throwing dice
- Drawing of lots

High risk of bias
- Sequence generated by a non-random component e.g.
  - odd or even date of
  - birth date (or day) of admission
  - hospital or clinic record number
- judgement of the clinician
- preference of the participant
- availability of the intervention
Allocation concealment (selection bias)

Low risk
- Central allocation (including telephone, web-based and pharmacy-controlled randomization)
- Sequentially numbered drug containers of identical appearance
- Sequentially numbered, opaque, sealed envelopes.

High risk
- Alternation or rotation
- Open random allocation schedule (e.g. a list of random numbers)
- Envelopes were unsealed or non-opaque
Performance bias

• Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.

• Blinding of participants, personnel and outcome assessors
Blinding (Performance bias)

Low risk of bias
• No blinding, but outcome and the outcome measurement are not likely to be influenced
• Blinding of participants and personnel
• Blinding of participants or personnel but outcome assessment unlikely to have been affected

High risk of bias
• No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding
• Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken
• No blinding
Attrition bias

• Systematic differences between groups in withdrawals from a study.

• **Attrition** refers to situations in which outcome data are not available

• **Exclusions** refer to situations in which some participants are omitted from reports of analyses, despite outcome data being available to the trialists.
Incomplete reporting (Attrition bias)

Low risk of bias
- No missing outcome data
- Reasons for missing outcome data unlikely to be related to true outcome
- Methodology ITT

High risk of bias
- Reason for missing outcome data likely to be related to true outcome,
- “As-treated’ analysis done with substantial departure of the intervention received from that assigned at randomization
Intention to treat (ITT)

• participants in trials should be analysed in the groups to which they were randomized, regardless of whether they received or adhered to the allocated intervention.

• 2 issues:
  – estimate the effects in practice
    • Not a subgroup who adhere to the intervention
    • “Per protocol” can overestimate effects
  – Loss to follow up
    • ITT ensures the outcome is still measured on these patients
Reporting bias

• systematic differences between reported and unreported findings.

• E.g publication bias, more likely to report significant differences between intervention groups than non-significant differences.
Selective outcome reporting (Reporting bias)

Low risk of bias

• The study protocol is available and all of the study’s pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way
• The study protocol is not available but it is clear that the published reports include all expected outcomes

High risk of bias

• Not all of the study’s pre-specified primary outcomes have been reported
• One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified
• One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);
• outcomes of interest in the review are reported
Other biases

• Trial designs
  – carry-over in cross-over trials
  – recruitment bias in cluster-randomized trials
    • E.g participants may know already which group they have been allocated to because everyone in that “cluster” gets the same intervention.
Cochrane risk of bias table

<table>
<thead>
<tr>
<th></th>
<th>Random sequence generation (selection bias)</th>
<th>Allocation concealment (selection bias)</th>
<th>Blinding (performance bias and detection bias)</th>
<th>Incomplete outcome data (attrition bias)</th>
<th>Selective reporting (reporting bias)</th>
<th>Other bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herrera Torres 2006</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
</tbody>
</table>
http://handbook.cochrane.org/front_page.htm
Decision-making quality vs. Amount of information.
# RRAMMbo tool map to Cochrane RoB

<table>
<thead>
<tr>
<th>RRAMMbo Domain</th>
<th>Question</th>
<th>Type of bias</th>
<th>Cochrane RoB_domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>Were the subjects representative of the target population?</td>
<td>Selection bias, Other sources of bias</td>
<td>Other sources of bias</td>
</tr>
<tr>
<td>Randomisation</td>
<td>How was randomisation carried out? Was allocation concealed?</td>
<td>Selection bias</td>
<td>Sequence generation, Allocation concealment</td>
</tr>
<tr>
<td>Allocation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>Were the groups equal at the start? And maintained through equal management and f/u?</td>
<td>Performance bias, Attrition bias</td>
<td>Incomplete outcome data, Blinding of participants, personnel and outcome assessors</td>
</tr>
<tr>
<td>Measurement-Blinding</td>
<td>Were the outcomes measured with blinded assessors/participants</td>
<td>Performance bias</td>
<td>Blinding of participants, personnel and outcome assessors</td>
</tr>
<tr>
<td>Objective outcomes (Measurement)</td>
<td>Were there differences in how outcomes were determined</td>
<td>Detection bias</td>
<td>Blinding of participants, personnel and outcome assessors. Other potential threats to validity</td>
</tr>
</tbody>
</table>
## Types of bias

<table>
<thead>
<tr>
<th>Type of bias</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection bias</td>
<td>Systematic differences between baseline characteristics of the groups that are compared.</td>
</tr>
<tr>
<td>Performance bias</td>
<td>Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Systematic differences between groups in withdrawals from a study</td>
</tr>
<tr>
<td>Detection bias</td>
<td>Systematic differences between groups in how outcomes are determined</td>
</tr>
<tr>
<td>Reporting bias</td>
<td>Systematic differences between reported and unreported findings</td>
</tr>
</tbody>
</table>