Brighton and Sussex NHS
Library and Knowledge Service

How to read a paper

ICU Journal club, 27 February 2018

Brighton and Sussex NHS Library and Knowledge Service is part of
Brighton and Sussex University Hospitals NHS Trust

“A lot of what is published is incorrect.”

- ‘…much of the scientific literature, perhaps half, may simply be untrue…’
  - small sample sizes
  - tiny effects
  - invalid exploratory analyses
  - flagrant conflicts of interest
  - obsession for pursuing fashionable trends of dubious importance
- ‘science has taken a turn towards darkness’

Horton, R **Offline: What is medicine's 5 sigma?**
Lancet 2015 April 11; 385 (9976): 1380

How do you assess a journal article?

Usual blood pressure, peripheral arterial disease, and vascular risk: cohort study of 4.2 million adults

Connor A Emdin,[1] Simon G Anderson,[1] Thomas Callender,[1] Nathalie Conrad,[1] Gholamreza Salimi-Khorshidi,[1] Hamid Mohseni,[1] Mark Woodward,[2,3] Kazem Rahimi[1,4]

[1]The George Institute for Global Health, Oxford Martin School, University of Oxford, Oxford OX1 3DB, UK

[2]The George Institute for Global Health, University of Sydney, Sydney, Australia

[3]Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA

[4]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Correspondence to: K Rahimi
kazem.rahimi@cardiov.ox.ac.uk

Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmj.h4865)

Cite this as: *BMJ* 2015;351:h4865
doi: 10.1136/bmj.h4865

Accepted: 26 August 2015

ABSTRACT

OBJECTIVES

To determine the subgroup specific associations between usual blood pressure and risk of peripheral arterial disease, and to examine the relation between peripheral arterial disease and a range of other types of vascular disease in a large contemporary cohort.

DESIGN

Cohort study.

SETTING

Linked electronic health records from 1990 to 2013 in the United Kingdom.

PARTICIPANTS

4 222 459 people aged 30-90 years, registered at a primary care practice for at least one year and with a blood pressure measurement.

MAIN OUTCOME MEASURES

Time to first diagnosis of new onset peripheral arterial disease and time to first diagnosis of 12 different vascular events.

RESULTS

A 20 mm Hg higher than usual systolic blood pressure was associated with a 63% higher risk of peripheral vascular event among those with peripheral arterial disease was chronic kidney disease (24.4% of initial events), followed by ischaemic heart disease (18.5% of initial events), heart failure (14.7%), and atrial fibrillation (13.2%). Overall estimates from this cohort were consistent with those derived from traditional studies when we pooled the findings in two meta-analyses.

CONCLUSIONS

Raised blood pressure is a strong risk factor for peripheral arterial disease in a range of patient subgroups. Furthermore, clinicians should be aware that those with established peripheral arterial disease are at an increased risk of a range of other vascular events, including chronic kidney disease, ischaemic heart disease, heart failure, atrial fibrillation, and stroke.

Introduction

In 2010 an estimated 202 million people globally had peripheral arterial disease.[1] The prevalence of the disease increases with age, from 1% of the population at age 40-49 years to 22.4% at age 80 or older.[2] In an

Critical appraisal – a definition

A process of assessing the validity, reliability and applicability of evidence

Validity – degree to which a study is likely to be true and free from bias (internal/external)
Reliability – would you get the same results if you repeatedly carried out the research?
Applicability – how well would the results apply to your own clinical practice?

Critical appraisal tools - CASP

Tools from the Critical Appraisal Skills Programme (CASP)

- Systematic Reviews
- Randomised Controlled Trials
- Qualitative Research Studies
- Cohort Studies
- Case-Control Studies
- Diagnostic Test Studies
- Economic Evaluation Studies

http://www.casp-uk.net/

Critical appraisal tools - CEBM

Centre for Evidence Based Medicine (CEBM)

- Systematic Review
- Diagnostic
- Prognosis
- RCT
- PICO

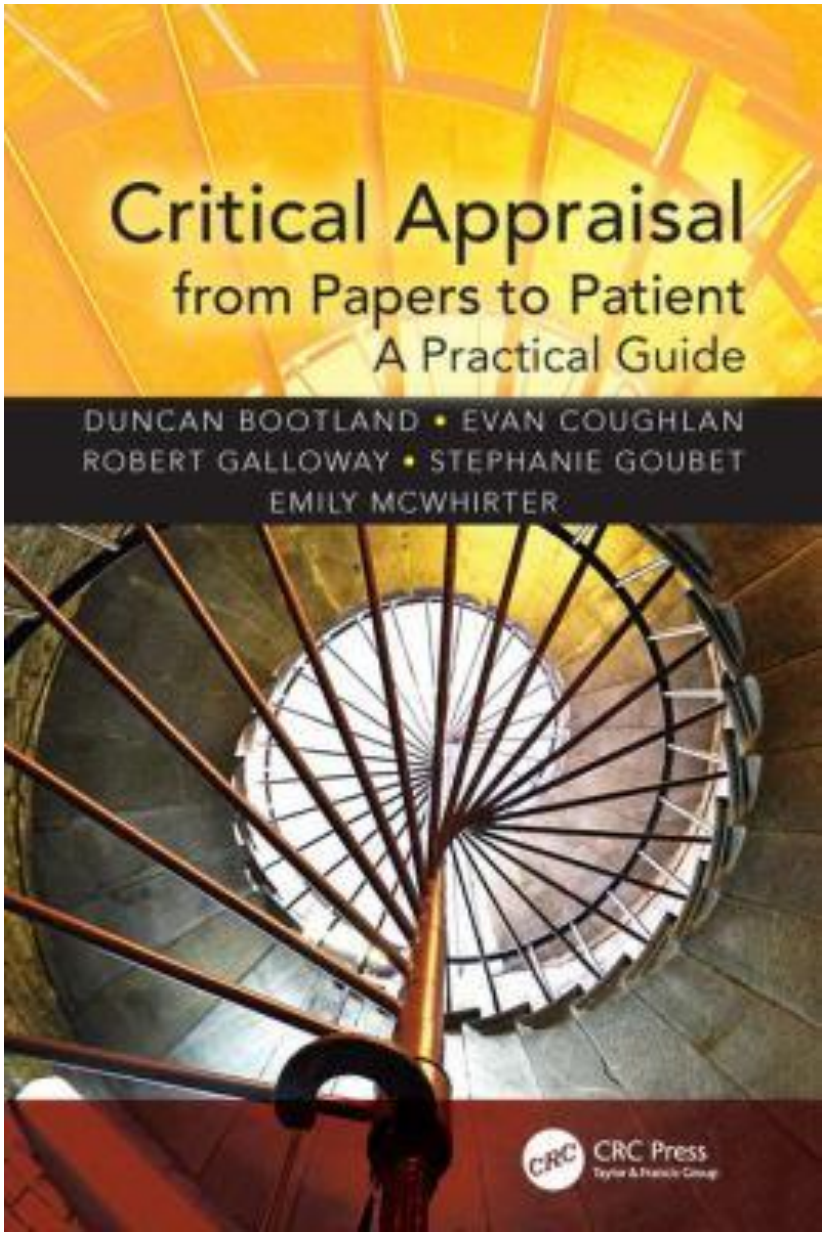
https://www.cebm.net/2014/06/critical-appraisal/

Critical appraisal tools - SIGN

Scottish Intercollegiate Guidelines Network (SIGN)

- Systematic Review and meta-analyses
- RCTs
- Cohort
- Case control
- Diagnostic
- Economic

http://www.sign.ac.uk/checklists-and-notes.html


Critical Appraisal
from Papers to Patient
A Practical Guide
DUNCAN BOOTLAND • EVAN COUGHLAN
ROBERT GALLOWAY • STEPHANIE GOUBET
EMILY MCWHIRTER
CRC Press

The framework

- Assess whether the research question was clear, sensible and relevant to your clinical situation
- Assess whether the most appropriate (and strongest possible) study design was used
- Identify opportunities for bias in the research method and judge whether these weaken the research conclusions
- Interpret the statistics to see if they support the conclusions drawn from the research
- Assess whether you can apply the research to your own clinical setting

Formulating a question

Formulating a question: PICO(T)

- Patient / Population:
  - Who is the question about?
- Intervention / Exposure:
  - What is being done or what is happening to the patient/population?
- Comparison:
  - What could be done instead of the intervention?
- Outcome:
  - How does the intervention affect the patient / population?
- Time frame

Other frameworks are available: ECLIPSE; SPICE

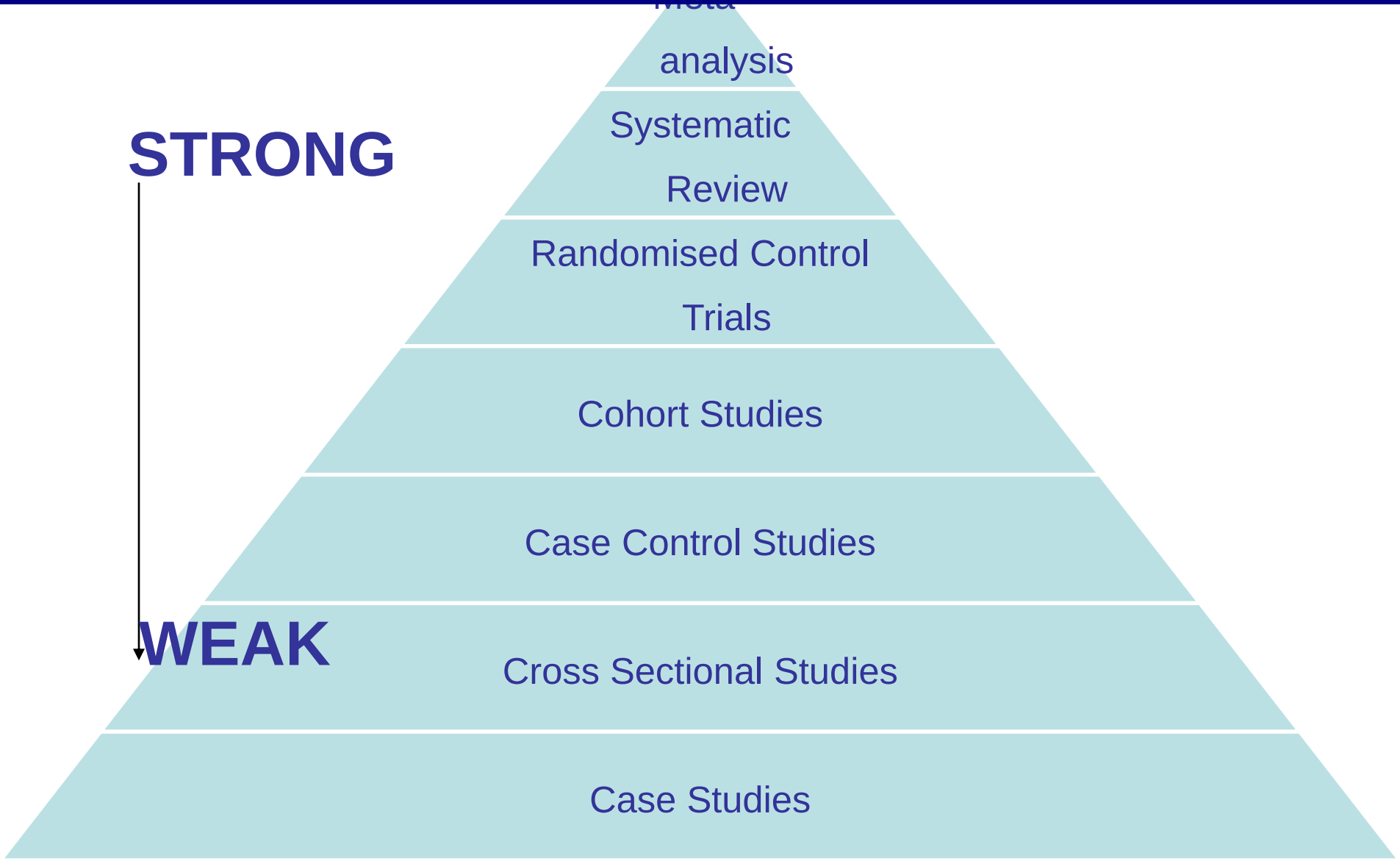
For a review of a number of frameworks see: Davies, K. Formulating the evidence based practice question: a review of the frameworks . *Evidence based library and information practice* 2011; 6(2): 75-80.

Some first questions to ask of the research question

- Were the participants, intervention, comparison and outcomes *clear?*
- Were the participants, intervention, comparison and outcomes *sensible?*
- Were all the outcomes clearly addressed in the results/discussion?

Study designs

Hierarchy of evidence



Study designs

Randomised Controlled Trial	Quantitative, prospective – subjects randomly assigned to one intervention or another.
Cross-sectional survey	Representative sample of the population is studied at a single point in time
Cohort Study	Prospective – takes a cohort of people that have received a particular exposure and a cohort that have not – follows them over time
Survey using Gold Standard	As cross-sectional survey above, but comparing test again gold standard
Qualitative (Interviews, Observations, etc)	Looks at issues in open-ended way and make sense of findings in terms of the meanings people bring to them

Study designs

Randomised Controlled Trial	Effectiveness of therapy
Cross-sectional survey	Incidence / prevalence
Cohort Study	Risk factors / prognosis
Survey using Gold Standard	Accuracy of a diagnostic test
Qualitative (Interviews, Observations, etc)	Attitudes & beliefs

Some historical examples: RCT

- **Lind J (1753) A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson**

Some historical examples: cohort study

- Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. BMJ. 1954;228:1451–1455

CEBM Levels of evidence 2

Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or *n*-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, *n*-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or *n*-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non -randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

*** Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.**

**** As always, a systematic review is generally better than an individual study.**

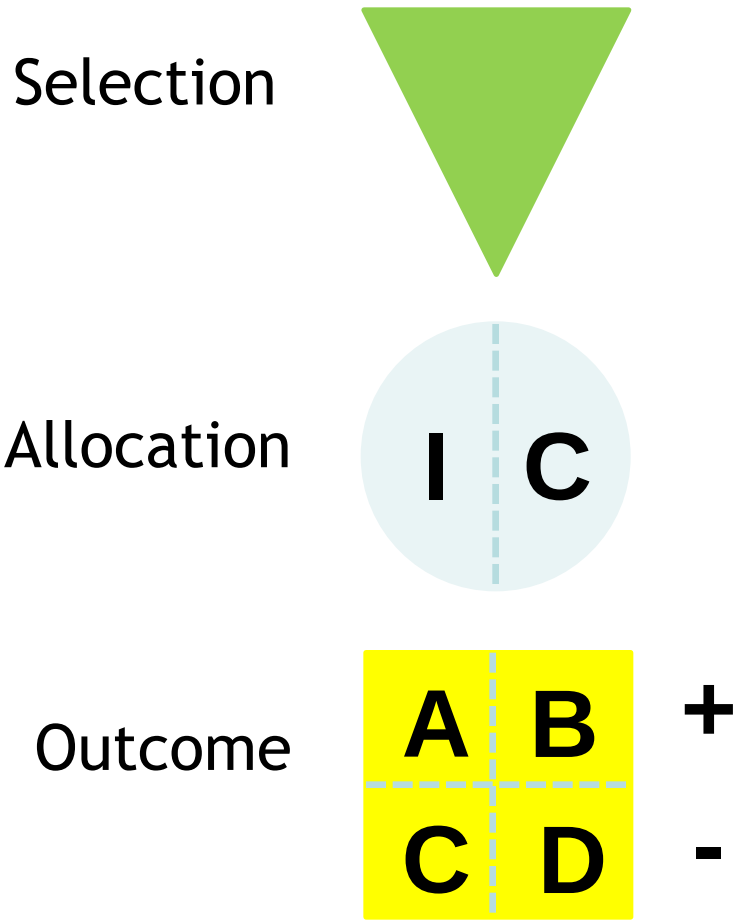
OCEBM Levels of Evidence Working Group*. "The Oxford 2011 Levels of Evidence". Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653

Recognising bias

What makes a study questionable?

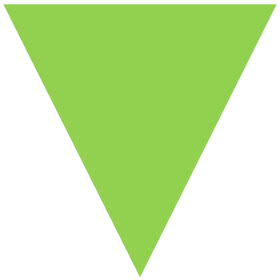
- Confounding Factors
  - Underlying factors that can affect the results, which are outside the control of the research team. For example, age, comorbidities, etc.
- Systematic Bias
  - Mistakes made by the investigators (not necessarily intentional) that result in a false conclusion. Systematic bias can also be introduced by participants and staff delivering supporting care.

Graphical appraisal tool of epidemiological studies



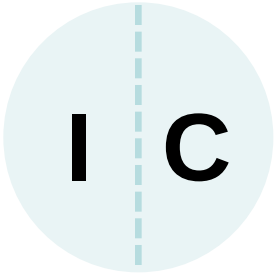
RAMMbo – remembering the areas where there is potential bias

Selection



Recruitment

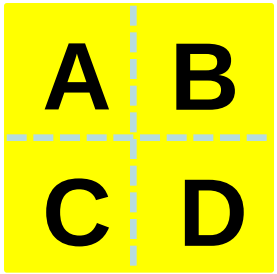
Allocation



Allocation

Maintenance

Outcome



+

-

Measurement
blinded
objective

Selection bias at *Recruitment*

- is it a representative sample?

- Occurs when the method of recruiting participants creates a non-representative group for study
  - Do the participants selected represent the population of interest?
  - Have they been selected randomly or sequentially?
  - Are the exclusion criteria relevant to the study?

RAMMbo

Some types of bias at recruitment

- Admission bias (Berkson)
- Diagnostic purity bias
- Incidence bias (Neyman)
- Membership bias
- Historical control

Selection bias at *Allocation*

– are the groups well-balanced?

- In general you can tell by the description of the randomisation process whether or not it was done correctly
  - It can happen that groups are not balanced in spite of randomisation
- The issue is not whether these differences are statistically significant. It is whether you feel they are large enough to affect the conclusion being drawn

RAMbo

Random sampling

- Stratified Randomisation
  - The number of participants allocated to intervention or control are balanced across strata, e.g. gender, age
- Random Permuted Blocks (A=intervention; B=control)
  - 20 – 40 yrs = ABAB; AABB; etc
  - 40 – 60 yrs = ABBA; ABAB; etc
  - 60 – 80 yrs = AABB; BBAA; etc
- Clustered Randomisation
  - Centres are randomised rather than individuals

RAMMbo

Concealed allocation

A process of masking, or hiding, which participant has been allocated to which arm of the trial.

Single blinding - the participant doesn't know, at the point when the decision was made, which arm of the trial s/he has been allocated to.

Double blinding - neither the trial administrator (i.e. the GP or clinician administering the treatment) nor the participant knows which arm of the trial the participant is being allocated to.

RAMMbo

Performance bias during *Maintenance*

- Occurs when what was measured happened because of the study itself rather than the intervention.
  - Ask whether participants have been treated the same way throughout the trial (apart from the intervention)
  - Have they been exposed to the same influences?

RAMbo

Attrition bias

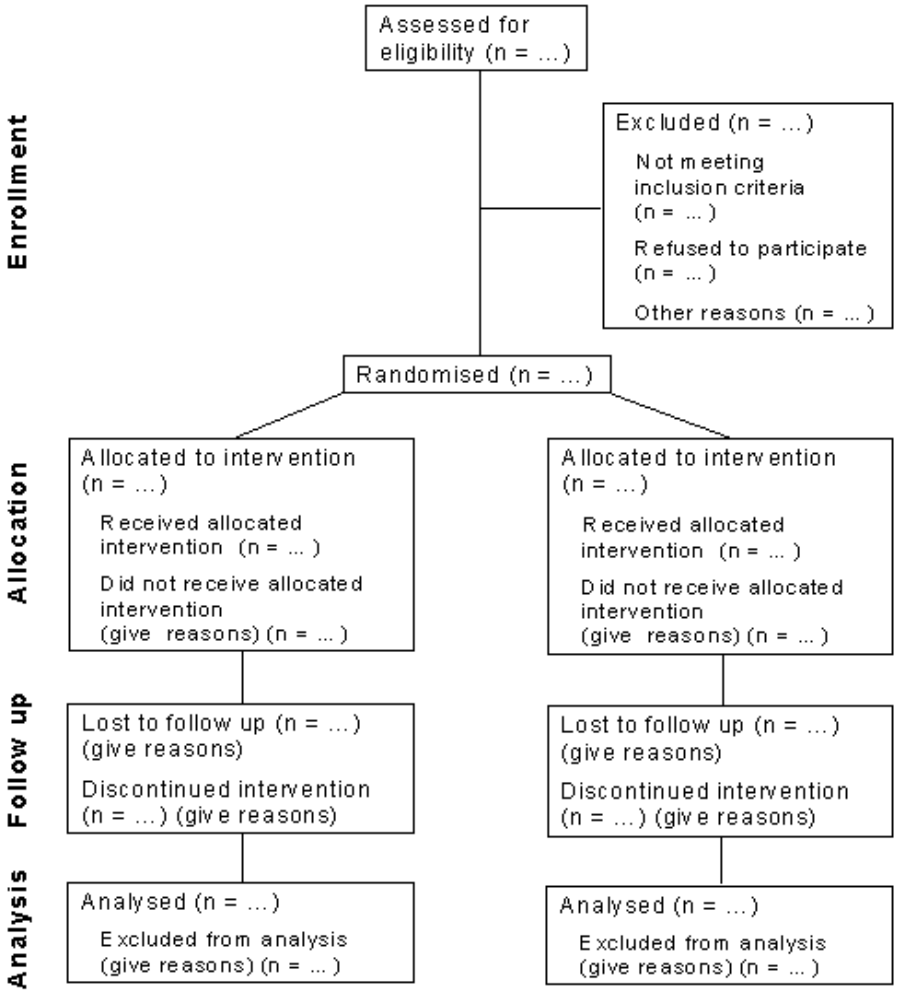
Occurs when there are important differences between the number of participants **lost to follow-up** in the comparison groups

Intention to Treat analysis says that data on *all* participants should be analysed with respect to the groups to which they were initially randomised.

Per protocol analysis – data on all participants is analysed with respect to the treatment they actually received.

RAMbo

Flow of patients through study



http://www.consort-statement.org/

Observer bias during *Measurement*

Occurs when those involved with the study* allow their knowledge of the study to affect the way observations are scored or recorded.

Blinding – single, double, triple

*This can mean: researchers, healthcare staff, or participants

RAMbo

Some types of observation bias

- Interviewer bias
- Recall bias
- Response bias

Blinding during *Measurement*

- Similar to concealed allocation but during the treatment and analysis elements of the trial
  - **Single blinding** - the participant doesn't know which arm of the trial s/he has been allocated to.
  - **Double blinding** - the trial clinician AND the participant are unaware of which arm of the trial the patient is in
  - **Triple blinding** – the statistician analysing the results does not know which arm is being analysed

RAMbo

Objectivity during *Measurement*

- Measurements should be both
  - Truly represent a measure of the outcomes of interest
  - Be as objective as is possible

RAMbo

Understanding the statistics

Types of statistics

Descriptive statistics - summarise the population and the results

Statistics for probability - describe the results as comparisons between the groups under study and demonstrate difference

Statistics for validity – describe the reliability of the study and how the results are applicable to others

Statistics for description

Descriptive statistics

Summarise a population and the results

Numerical – where a value can fall at any point in a range (e.g. weight):

1) *Discrete – number of people*
2) *Continuous – height, weight*

2) *Categorical* (nominal/ordinal) – where a value is selected from specific options.Nominal can be binary (M/F) or multi-categorical (eg colour)

Some measurements can fall into either – BMI (e.g. 28, or 'overweight')

Different techniques are used to summarise each type of data.

Statistics for probability

Common terms used in comparing study groups

- Absolute risk
- Relative risk
- Risk ratio
- Hazard ratio
- Odds ratio

All of these compare how often the event that happens in Group X happens in Group Y

Risk and number needed to treat

Absolute Risk Reduction: CER – EER

or

Absolute Risk Reduction: RRR x CER

Number Needed to Treat: 1 ÷ ARR (or 100 ÷ ARR, if ARR expressed as a percentage)

[Number of people to treat with an intervention to prevent one outcome]

Statistics for validity

Confidence intervals

	Vitamin D and calcium	Calcium alone
% of people who fell	16%	28%

Relative Risk Reduction = 0.43

Statistics show that we can be 95% sure that the *true* RRR is between 0.29 and 1.06

i.e. RRR = 0.43 (95% CI, 0.29-1.06)

Confidence intervals

We use the information we have about the average and spread of our sample data…

… or information on the proportions in our sample…

… combined with the sample size…

… to calculate a range within which we expect the true value to be.

Does the intervention work?

Actual effect of treatment \ Results of study	Study finds a difference between groups	Study finds no difference between groups
Groups are really the same	Type I Error = False Positive [p-value]	No Error
Groups are really different	No Error	Type II Error = False Negative [Power]

Does the intervention work?

Actual effect of treatment \ Results of study	Study finds that Vitamin D **does** reduce falls	Study finds that Vitamin D **does not** reduce falls
Vitamin D **does not** reduce falls	Type I Error = False Positive [p-value]	No Error
Vitamin D **does** reduce falls	No Error	Type II Error = False Negative [Power]

P-values

p-value =

The probability that a difference observed between two groups has occurred by chance

If we're trying to show a real difference between X and Y we want the p-value to be small

Ideally p less than 0.05 or 5%

NB. The p-value is also known as α (alpha)

P-values

- If $p < 0.05$ there is a high chance that the difference between the groups is *real.*
- If $p > 0.05$ then either:
  - A) There is no real difference between the groups OR
  - B) There *is* a difference but the study has failed to detect it because of random chance
- To determine which of A or B is more likely, look at the Power

Power

- The Power is the chance that if a difference exists the study will detect it.
- A small sample may lead you to conclude that there isn't a difference between two groups, when actually there is.
- More participants = higher Power = good!
- Ideally more than 0.8

Power

In the case where $p > 0.05$

- If Power > 0.8 then: there probably is no difference between the groups.
- If Power < 0.8 then: it is possible that the researchers have failed to detect a difference between the groups.

References

Books

- Ajetunmobi, O. (2002). *Making sense of critical appraisal.* Arnold.
- Bootland, D, Coughlan E, Galloway R et al (2017) Boca Raton: CRC Press
- Gosall NK, Gosall GS (2015) *The doctor's guide to critical appraisal* 4th ed Knutsford: PasTest
- Greenhalgh, T. (2014). *How to read a paper.* 5th ed. London: Wiley-Blackwell / BMJ Publishing.
- Guyatt, G. (2015). *Users' guides to the medical literature.* 3rd ed McGraw-Hill Education
- Harris, M. and G. Taylor (2014). *Medical statistics made easy*. 3rd ed. Banbury: Scion.
- Kranzler, J H (2011). *Statistics for the terrified.* 5th ed. London: Pearson.
- Petrie, A and C Sabin (2009). *Medical statistics at a glance* 3rd ed. Chichester: Wiley.
- Peat, J, B Barton and E Elliott (2008). *Statistics workbook for evidence-based health care* Chichester: Wiley.

Websites:

- Centre for Evidence Based Medicine: www.cebm.net/
- Critical Appraisal Skills Programme: www.casp-uk.net/
- www.healthknowledge.org.uk/interactive-learning/finding-and-appraising-the-evidence
- Prisma statement: www.prisma-statement.org/
- Consort statement: www.consort-statement.org/