

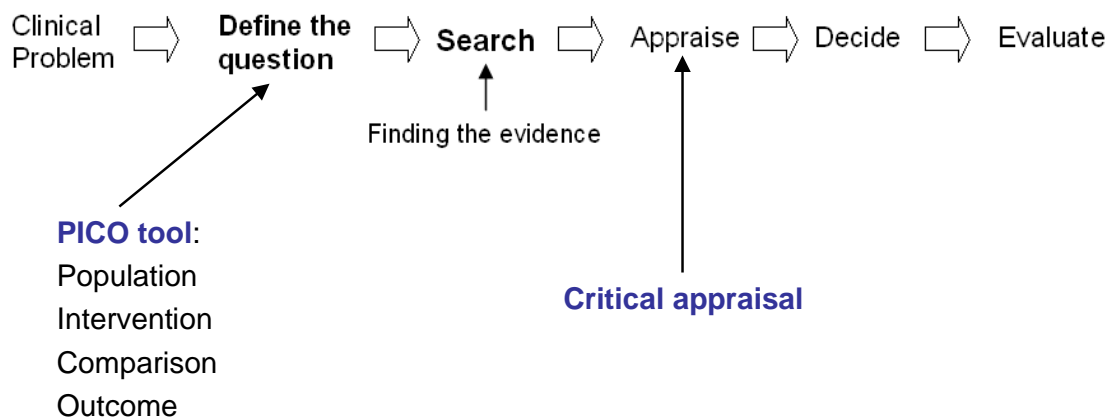


1. Introduction to critical appraisal

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context. (Burls 2009)

Critical appraisal is an important element of **evidence-based medicine**. The **five steps** of evidence-based medicine are:

1. asking **answerable questions**, i.e. formulating questions into a format whereby you can interrogate the medical literature and hopefully find an answer - to do this, you may use the **PICO tool**, which helps to break down the query into Population, Intervention, Comparison, Outcome;
2. you then need to **search** for the evidence - if you can find a pre-appraised resource, you can miss out the next step;
3. the next step is **critical appraisal** of your results;
4. you then decide what **action** to take from your findings;
5. finally, you **evaluate** your new or amended practice.



Critical appraisal is essential to:

- combat **information overload**;
- identify papers that are **clinically relevant**;
- **Continuing Professional Development (CPD)** - critical appraisal is a requirement for the evidence based medicine component of many membership exams.

2. Location and selection of studies

2.1. *Bad science*

We often come across news articles making unjustified scientific/medical claims. For example, in June 2008 the Sunday Express published an article about the link between suicides and phone masts:

The spate of deaths among young people in Britain's suicide capital could be linked to radio waves from dozens of mobile phone transmitter masts near the victims' homes.

Dr Roger Coghill, who sits on a Government advisory committee on mobile radiation, has discovered that all 22 youngsters who have killed themselves in Bridgend, South Wales, over the past 18 months lived far closer than average to a mast. (Johnston 2008)

Ben Goldacre, a medical doctor and former author of the weekly **Bad Science** column in the Guardian, investigated the claim made by the Sunday Express article and found out the following:

I contacted Dr Coghill, since his work is now a matter of great public concern, and it is vital his evidence can be properly assessed. He was unable to give me the data. No paper has been published. He himself would not describe the work as a "study". There are no statistics presented on it, and I cannot see the raw figures. In fact Dr Coghill tells me he has lost the figures. Despite its potentially massive public health importance, Dr Coghill is sadly unable to make his material assessable. (Goldacre 2008)

2.2. *Behind the headlines*

The article about the link between suicides and phone masts is an example of the way in which 'bad science' can make it to the headlines. Sometimes, however, science/health stories found in the news are genuinely based on valid studies, but jump to wrong conclusions by failing to consider some important aspects, such as the study design and the level of evidence of the original research.

For instance, in July 2008 an article was published on the Daily Mail claiming that there is a link between vegetarian diet and infertility (Daily Mail Reporter 2008). The article was based on a cross-sectional study on soy food intake and semen quality published in the medical journal Human Reproduction (Chavarro et al. 2008). **Behind the Headlines**, a NHS service providing an unbiased daily analysis of the science behind the health stories that make the news, issued the following comment:

The Daily Mail today reports on, "Why a vegetarian diet may leave a man less fertile." It said research has found that eating tofu can significantly lower your sperm count.

The study behind this news had some limitations: it was small, and mainly looked at overweight or obese men who had presented to a fertility clinic. It focused only on soy (soya) intake, and the Daily Mail's claim that there is a causal link between eating a 'vegetarian diet' and reduced fertility is misleading. (NHS Knowledge Service 2008)

2.3. *Bias in the location and selection of studies*

Perhaps it is not surprising that the study on soy and infertility received some publicity - but if the study had not obtained positive results, would it have been published - and quoted in the news?

When reviewing the literature published in scientific/medical journals, we should consider that papers with **significant positive results** are more likely to be:

- submitted and accepted for publication (**publication bias**);
- published in a major journal written in English (**Tower of Babel bias**);
- published in a journal indexed in a literature database, especially in less developed countries (**database bias**);
- cited by other authors (**citation bias**);
- published repeatedly (**multiple publication bias**);
- ... and quoted by newspapers!
(Egger & Smith 1998; Gregoire, Derderian, & Le Lorier 1995)

3. Study design

The following lists summarise the most common types of study design found in the medical literature.

3.1. Qualitative studies

Qualitative studies explore and understand people's beliefs, experiences, attitudes, behaviour and interactions. They generate non-numerical data. Examples of qualitative studies:

- **Document** - study of documentary accounts of events, such as meetings;
- **Passive observation** - systematic watching of behaviour and talk in natural occurring settings;
- **Participant observation** - observation in which the researcher also occupies a role or part in the setting, in addition to observing;
- **In depth interview** - face to face conversation with the purpose of exploring issues or topics in detail. Does not use preset questions, but is shaped by a defined set of topics;
- **Focus group** - method of group interview which explicitly includes and uses the group interaction to generate data. (Greenhalgh 2001)

3.2. Quantitative studies

Quantitative studies generate numerical data or data that can be converted into numbers. Examples of quantitative studies:

- **Case report** - report on a single patient;
- **Case series** - report on a series of patients (no control group);
- **Case control study** - identifies patients with a particular outcome (cases) and control patients without the outcome. Looks back and explores exposures and possible links to outcome. Very useful in causation research;
- **Cohort study** - identifies two groups (cohorts) of patients one which received the exposure of interest, and one which did not. Follows these cohorts forward for the outcome of interest. Very useful in causation as well as prognosis research.
(Bandolier 2004)

Key quantitative studies:

- **Randomized Controlled Trial (RCT)** - a clinical trial in which participants are randomly allocated to a test treatment and a control; involves concurrent enrolment and follow-up of both groups; gold standard in testing the efficacy of an intervention (therapy/prevention);
- **Systematic review** - identifies and critically appraises all research on a specific topic, and combines valid studies; increasingly important in evidence based medicine; different from **review article** (which is a summary of more than one paper on a specific topic, and which may or may not be comprehensive);

- **Meta-analysis** - a systematic review that uses quantitative methods to summarise the results. (Bandolier 2004; NCBI 2010)

The following diagram shows a model for the organisation of some quantitative studies. Different types of studies are located at different levels of the **hierarchy of evidence**. All types of studies may be found published in journals, with the exception of the top two levels.



Adapted from (Haynes 2006).

There are also **other types** of quantitative studies, such as:

- **Cross-sectional survey** - the observation of a defined population at a single point in time or time interval. Exposure and outcome are determined simultaneously. Gold standard in diagnosis and screening research;
- **Decision analysis** - uses the results of primary studies to generate probability trees to be used in making choices about clinical management or resource allocation;
- **Economic analysis** - uses the results of primary studies to say whether a particular course of action is a good use of resources. (Bandolier 2004; Greenhalgh 2001)

3.3. Critical appraisal of different study designs

To critically appraise a journal article, you would have to start by assessing the research methods used in the study. This is done using **checklists** which are specific to the study design. The following checklists are commonly used:

- **CASP** <http://www.casp-uk.net/checklists>;
- **SIGN** <http://www.sign.ac.uk/methodology/checklists.html>;
- **CEBMH** http://cebmh.warne.ox.ac.uk/cebmh/education_critical_appraisal.htm.

4. Randomised Controlled Trials (RCTs)

4.1. Mechanisms to control bias in RCTs

RCTs **control bias** by randomisation and blinding.

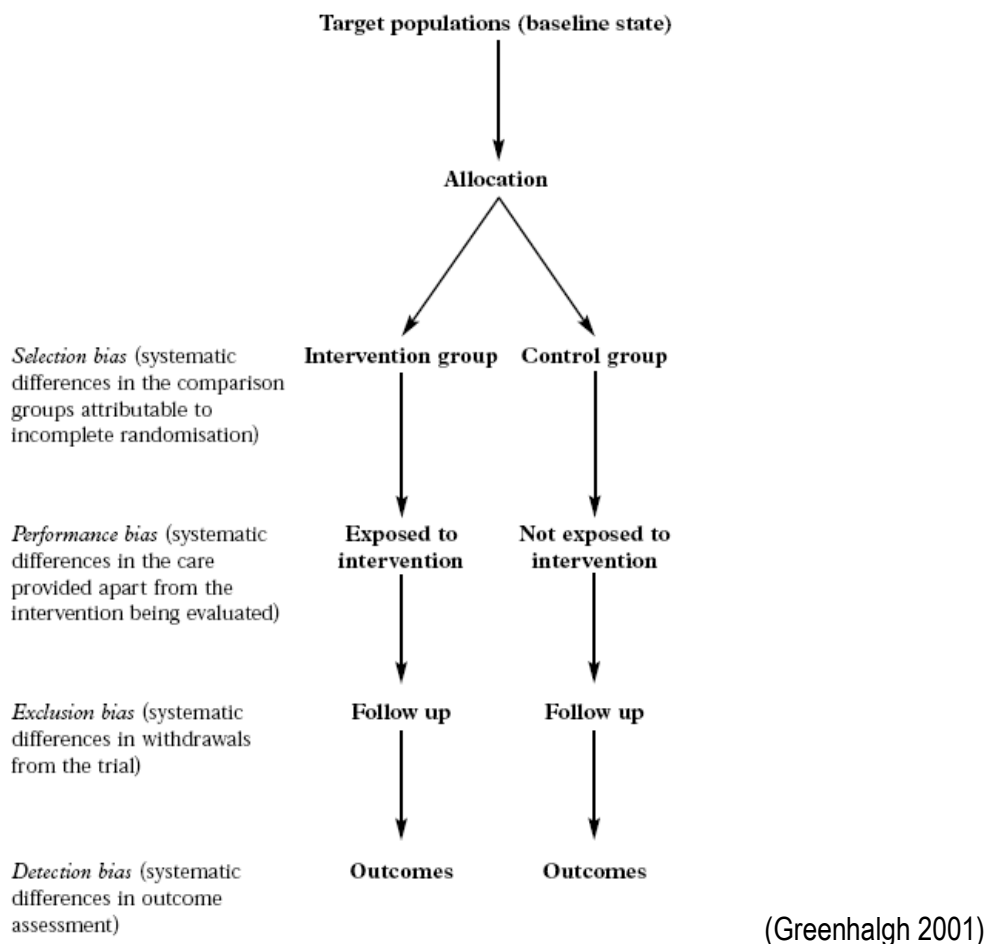
Randomisation indicates that participants are randomly allocated to treatment or control group.

- Acceptable methods of randomisation include random numbers, either from tables or computer-generated (for more details see Schulz & Grimes 2002).
- Unacceptable methods include last digit of date of birth, date seen in clinic etc. (for more details see Stewart & Parmar 1996).
- **Stratified randomisation** is often used to avoid **confounding** factors, i.e. to ensure equal distribution of participants with a characteristic thought to affect prognosis or response.

Blinding means masking who is getting treatment and control.

- Single blinding: participants do not know.
- Double blinding: neither the participants nor those giving the intervention know.
- Triple blinding: statisticians doing the analysis also do not know.

The following diagram illustrates the sources of bias in RCTs:



4.2. Advantages and disadvantages of RCTs

Advantages:

- allow for rigorous evaluation of a single variable;
- potentially eradicate bias;
- allow for meta-analysis.

Disadvantages:

- expensive;
- time consuming;
- **ethically problematic** at times - a trial is sometimes stopped early if dramatic effects are seen.

4.3. Preliminary statistical concepts in RCTs

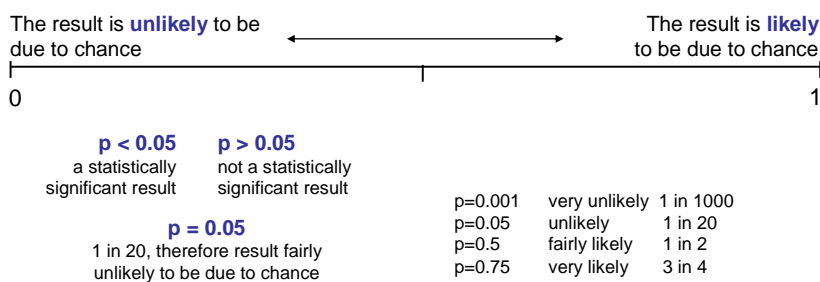
Baseline characteristics - both the control and the intervention group should be broadly similar in factors like age, sex distribution and level of illness.

Sample size calculation (Power calculation) - a trial should be big enough to have a high chance of detecting a worthwhile effect if it exists. Statisticians can work out before the trial begins how large the sample size should be in order to have a good chance of detecting a true difference between the intervention and control groups (Greenhalgh 2001). Standard power: 80%.

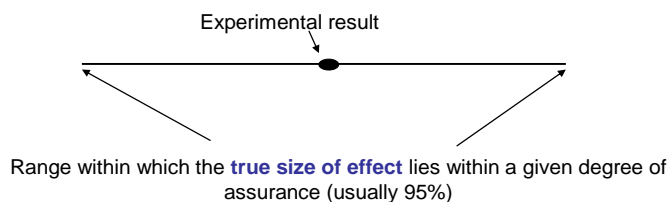
Intention to treat - all data on participants including those who withdraw from the trial should be analysed. Failure to do so may lead to underestimation/overestimation of results (Hollis & Campbell 1999).

4.4. Presenting the results of RCTs

P-value - the p-value refers to the probability that any particular outcome would have arisen by chance. A p-value of less than 1 in 20 ($p < 0.05$) is statistically significant.



Confidence interval - the same trial repeated hundreds of times would not yield the same results every time. But on average the results would be within a certain range. A 95% confidence interval means that there is a 95% chance that the true size of effect will lie within this range.



4.5. Quantifying the risk of benefit/harm in RCTs

Experimental Event Rate (EER) - in the treatment group, number of patients with outcome divided by total number of patients.

Control Event Rate (CER) - in the control group, number of patients with outcome divided by total number of patients.

Relative Risk or Risk Ratio (RR) - the risk of the outcome occurring in the intervention group compared with the control group.

$$RR = EER / CER$$

Absolute Risk Reduction or increase (ARR) - absolute amount by which the intervention reduces (or increases) the risk of outcome.

$$ARR = CER - EER$$

Relative Risk Reduction or increase (RRR) - amount by which the risk of outcome is reduced (or increased) in the intervention group compared with the control group.

$$RRR = ARR / CER$$

Odds of outcome - in each patient group, the number of patients with an outcome divided by the number of patients without the outcome.

Odds ratio - odds of outcome in treatment group divided by odds of outcome in control group.

If the outcome is negative, an effective treatment will have an odds ratio <1 ;

If the outcome is positive, an effective treatment will have an odds ratio >1 .

(In case control studies, the odds ratio refers to the odds in favour of exposure to a particular factor in cases divided by the odds in favour of exposure in controls).

Number needed to treat (NNT) - how many patients need to have the intervention in order to prevent one person having the unwanted outcome.

$$NNT = 1 / ARR$$

Ideal NNT=1;

The higher the NNT, the less effective the treatment.

4.6. Critical appraisal of RCTs

Factors to look for:

- allocation (randomisation, stratification, confounders);
- blinding;
- follow up of participants (intention to treat);
- data collection (bias);
- sample size (power calculation);
- presentation of results (clear, precise);
- applicability to local population.

5. Systematic reviews

5.1. Mechanisms to control bias in systematic reviews

Systematic reviews provide an overview of all primary studies on a topic and try to obtain an overall picture of the results.

To avoid bias, systematic reviews must:

- contain a statement of **objectives**, **materials** and **methods**;
- follow an explicit and reproducible **methodology** (Greenhalgh 2001).

In a systematic review, all the primary studies identified are critically appraised and only the best ones are selected. A **meta-analysis** (i.e. a statistical analysis) of the results from selected studies may be included.

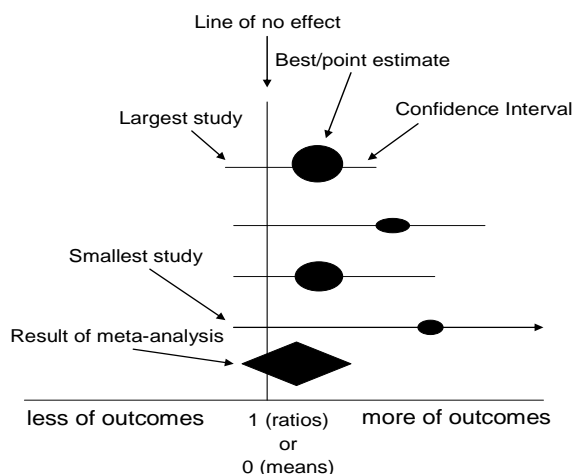
5.2. Blobbogram/Forrest plot

A **blobbogram** or **forest plot** is a graphical display used to present the result of a meta-analysis.

Selected studies must be tested for **homogeneity**, which should be >50%. A quick way to check for homogeneity is to look at the confidence intervals for each study - if they don't overlap, the studies are likely to be heterogeneous. More rigorous tests of homogeneity include χ^2 .

If studies are homogeneous, a **fixed-effect model** is normally used in the meta-analysis. This means that results are only interpreted within the populations/samples in the included studies.

If studies are heterogeneous, a **random-effects model** is used. This means that results are interpreted across the wider population. A different underlying effect is assumed for each study and an additional source of variation is added to the model.



5.3. Advantages and disadvantages of systematic reviews

Advantages:

- allow for rigorous pooling of results;
- may increase overall confidence from small studies;
- potentially eradicate bias;
- may be updated if new evidence becomes available;
- may have the final say on a clinical query;
- may identify areas where more research is needed.

Disadvantages:

- expensive;
- time consuming;
- may be affected by publication bias - a test called **Funnel Plot** can be used to test for publication bias;
- normally summarise evidence up to two years before (due to the time required for the execution of the systematic review).

5.4. Critical appraisal of systematic reviews

Factors to look for:

- literature search (did it include published and unpublished materials as well as non-English language studies? Was personal contact with experts sought?);
- quality-control of studies included (type of study; scoring system used to rate studies; analysis performed by at least two experts);
- homogeneity of studies;
- presentation of results (clear, precise);
- applicability to local population.

6. Where next?

You could set up an evidence-based journal club:

- choose a topic of interest in your group;
- one person performs a literature search and finds a paper to bring to the meeting;
- the paper is presented in the meeting, and the literature search is also explained;
- appraise the paper as a group.

7. Further information, support and training

7.1. Further reading

A number of books and journal articles have been written on critical appraisal. A good summary is provided by Guyatt & American Medical Association (2008):

Guyatt, G. & American Medical Association. 2008, Users' guides to the medical literature: a manual for evidence-based clinical practice / edited by Gordon Guyatt ... [et. al.] McGraw Hill Medical; JAMA & Archives Journals.

Hollands, H. & Kertes, P. J. 2010, "Measuring the size of a treatment effect: relative risk reduction, absolute risk reduction, and number needed to treat", Evidence-Based Ophthalmology, vol. 11, no. 4, pp. 190-194.

7.2. Online resources

LIBRARY CRITICAL APPRAISAL PAGE –

<https://www.ucl.ac.uk/child-health/support-services/library/training-and-clinical-information-support/critical-appraisal-resources>

AGREE - <http://www.agreetrust.org/>

AGREE is an international collaboration of researchers and policy makers who seek to improve the quality and effectiveness of clinical practice guidelines by establishing a shared framework for their development, reporting and assessment. The website contains the 'Agree Instrument' which provides a framework for assessing the quality of clinical practice guidelines.

Alberta University Evidence Based Medicine Toolkit - <http://www.ebm.med.ualberta.ca/>

This is a collection of tools for identifying, assessing and applying relevant evidence for better health care decision-making. The appraisal tools are adapted from the Users' Guides series prepared by the Evidence Based Medicine Working Group and originally published in JAMA. It includes a glossary.

CASP - <http://www.casp-uk.net>

The Critical Appraisal Skills Programme (CASP) is a programme within Workforce Development by Solutions for Public Health (SPH), a not-for-profit NHS public health organisation. CASP aims to enable individuals to develop the skills to find and make sense of research evidence. The website gives access to critical appraisal checklists which guide the appraisal of different types of study.

CATwalk - <http://guides.library.ualberta.ca/catwalk>

This site has been designed to assist University of Alberta medical residents in the process of completing a Critically Appraised Topic (CAT).

CEBMH - http://cebmh.warne.ox.ac.uk/cebmh/education_critical_appraisal.htm

Another useful source of guidelines by the Centre for Evidence-Based Mental Health.

Centre for Evidence Based Medicine - <http://www.cebm.net>

The Centre for Evidence Based Medicine is the first of several centres around the country whose broad aim is to promote evidence-based health care and provide support and resources to anyone who wants to make use of them. It includes a wide range of EBM resources including critical appraisal tools.

Dr Chris Cates' EBM Website - www.nntonline.net

Provides help with statistics.

CLIST Resources for Critical Appraisal - <http://www.londonlinks.nhs.uk/groups/clinical-librarians-information-skills-trainers-group/trainers-toolkit/resources-for-critical-appraisal>

Provides links to critical appraisal resource pages maintained by a selection of UK healthcare libraries.

The Little Handbook of Statistical Practice - <http://www.tufts.edu/~gdallal/LHSP.HTM>

Provides help with statistics.

How to read a paper - <http://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper>

Links to the series of articles that make up the book 'How to read a paper'. The articles are available online free of charge from the BMJ website.

New Zealand Guidelines Group – <https://www.health.govt.nz/about-ministry/ministry-health-websites/new-zealand-guidelines-group>

The NZGG exists to promote effective delivery of health and disability services, based on evidence. The webpage contains critical appraisal tools and guidance (under 'Evidence Resources').

SCHARR - <https://www.sheffield.ac.uk/scharr/sections/ir>

The School of Health and Related Research at the University of Sheffield links to useful web resources.

SIGN - <http://www.sign.ac.uk/guidelines/fulltext/50/>

The guideline developer's handbook by the Scottish Intercollegiate Guidelines Network is a useful source of guidelines.

University of Glasgow General Practice & Primary Care – Evidence Based Practice -

<http://www.gla.ac.uk/departments/generalpracticeprimarycare/ebp/#d.en.19511>

Useful collection of materials to help develop and practise the skills of critical appraisal, including checklists, 'jargon busters' to explain the terminology and worked examples. A list of other evidence-based practice sites is also included.

7.3. The Library

The Library offers training and support in a range of information skills on a group or individual basis. We can train you in the Library or in your workplace. Please contact:

Please see our website for more information:

<https://www.ucl.ac.uk/child-health/support-services/library/training-and-clinical-information-support>

References

Bandolier. Glossary index. Bandolier . 2004.

Web/URL: <http://www.bandolier.org.uk/glossary.html> Accessed March 2017.

Chavarro, J. E., Toth, T. L., Sadio, S. M., & Hauser, R. 2008, "Soy food and isoflavone intake in relation to semen quality parameters among men from an infertility clinic", *Human Reproduction*, vol. 23, no. 11, pp. 2584-2590.

Daily Mail Reporter. Why a vegetarian diet may leave a man less fertile. Daily Mail , 5. 24-7-2008.

Web/URL: <http://www.dailymail.co.uk/health/article-1038081/Why-vegetarian-diet-leave-man-fertile.html#ixzz15X6U5fn1> Accessed March 2017.

Egger, M. & Smith, G. D. 1998, "Bias in location and selection of studies", *BMJ*, vol. 316, no. 7124, pp. 61-66.

Goldacre, B. Bad science: Suicides, Aids, and a masts campaigner. The Guardian , 12. 28-6-2008.

Web/URL: <http://www.guardian.co.uk/commentisfree/2008/jun/28/sciencenews.mobilephones> Accessed March 2017

Greenhalgh, T. 2001, *How to read a paper : the basics of evidence based medicine / Trisha Greenhalgh* BMJ Books.

Gregoire, G., Derderian, F., & Le Lorier, J. 1995, "Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias?", *Journal of Clinical Epidemiology*, vol. 48, no. 1, pp. 159-163.

Guyatt, G. & American Medical Association. 2008, *Users' guides to the medical literature : a manual for evidence-based clinical practice / edited by Gordon Guyatt ... [et. al.]* McGraw Hill Medical ; JAMA & Archives Journals.

Haynes, R. B. 2006, "Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions", *Evidence-Based Medicine*, vol. 11, no. 6, pp. 162-164.

Hollands, H. & Kertes, P. J. 2010, "Measuring the size of a treatment effect: relative risk reduction, absolute risk reduction, and number needed to treat", *Evidence-Based Ophthalmology*, vol. 11, no. 4, pp. 190-194.

Hollis, S. & Campbell, F. 1999, "What is meant by intention to treat analysis? Survey of published randomised controlled trials", *BMJ*, vol. 319, no. 7211, pp. 670-674.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & Mcquay, H. J. 1996, "Assessing the quality of reports of randomized clinical trials: Is blinding necessary?", *Controlled Clinical Trials*, vol. 17, no. 1, pp. 1-12.

Johnston, L. Suicides 'linked to phone masts'. Sunday Express , 1. 22-6-2008.

Web/URL: [http://www.express.co.uk/posts/view/49330/Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts](http://www.express.co.uk/posts/view/49330/Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts-Suicides-linked-to-phone-masts) Accessed March 2017.

NCBI. MeSH. MeSH Database . 2010.

Web/URL: <http://www.ncbi.nlm.nih.gov/mesh> Accessed March 2017.

NHS Knowledge Service. Soya-based food and male fertility. Behind the Headlines . 24-7-2008.

Web/URL: www.nhs.uk/news/2008/07/July/Pages/Soyabasedfoodandmalefertility.aspx Accessed March 2017.

Schulz, K. F. & Grimes, D. A. 2002, "Generation of allocation sequences in randomised trials: chance, not choice", *Lancet*, vol. 359, no. 9305, pp. 515-519.

Sharma, S. 2010, "Levels of evidence", *Evidence-Based Ophthalmology*, vol. 11, no. 2, pp. 73-74.

Stewart, L. A. & Parmar, M. K. 1996, "Bias in the analysis and reporting of randomized controlled trials", *International Journal of Technology Assessment in Health Care*, vol. 12, no. 2, pp. 264-275.