**GLOSSARY of Research Terminology** (From MacPherson H, Hammerschlag R, Lewith G, Schnyer R (eds). *Acupuncture Research* Edinburgh, Churchill Livingstone, 2007:xix-xxvi)

#### Allocation concealment: in controlled clinical trials, the concealment of the allocation to groups is required in order to be confident that researchers or those providing treatment do not (unconsciously or otherwise) influence who gets allocated to which group. By helping prevent selection bias, this procedure reinforces the benefits of randomisation.

Audit: In clinical practice an audit is a process used to identify whether certain standards are met, such as sending out letters to referring physicians, or reporting of side effects. An audit is often conducted as part of an audit cycle in which changes are made to practice and another audit conducted after an agreed time. A less common use of the term audit is in the context of collecting data, usually retrospectively, on patients attending a particular clinic.

Blinding/blinded: randomised controlled trials are fully blinded if all the people involved are unaware of the treatment group to which trial participants are allocated until after the analysis of the results. It is not feasible for acupuncturists to be blinded in controlled trials, as they will be aware of the treatment that they are delivering. In explanatory randomised controlled trials with a sham control group, efforts are made to keep the patients blind to whether they receiving the real or the sham acupuncture intervention. Pragmatic randomised controlled trials are usually “open” trials with the patients aware of the treatment they are receiving. It is possible for the statistician to be blind to allocation when conducting the analysis. The terms “single blind” and “double blind” are not used consistently in the literature.

Case study: this is a case report based on clinical observation of one patient, often with some assessment of outcome.

Case series: repeated observations of more than one patient, often woth data collected to identify whether there is an association between treatment and improvement. This also may help establish whether particular patients or treatments are associated with a relatively better (or poorer) outcome. There is no control group.

Cohort study: in a cohort study a group of people are identified who have something in common – for example, a new diagnosis of hepatitis C – and followed over time to see the outcome, without any treatment or other intervention. The term is not usually used for a study of outcome from treatment, for which a case series is usually the correct term.

Complex intervention: an intervention can be defined as complex when there are multiple components where it is not clear how to separate out the “active” components of treatment from those that are inactive. It is possible that the sum of all the components are synergistic, interacting in such a way that the whole can be expected to be more effective than the sum of the parts.

Confidence interval (CI): Confidence intervals present the range of likely effects from an intervention. For example, a 95% confidence interval provides a range that will include 95% of results from studies of the same size and design in the same population. This can be thought of as there being a 95% chance that the true (but not exactly known) size of the effect falls within the confidence interval.

Controlled trial: a comparison between two or more different treatment groups, for example one group receiving acupuncture and another group acting as a control, who might receive usual care, some other treatment, or sham acupuncture. The term “controlled trial” is often used for trials without random allocation to groups, in which case there is potential selection bias, that is differentials between groups that may mean that they are not equivalent. Non-randomised controlled will have weaker findings due to suspected bias. When the method of allocation is by random selection, the study is referred to as a randomised controlled trial (RCT; see below).

Correlation: a correlation is a quantification of the strength of an association between two variables. A correlation will not indicate whether there is causation, see mechanism below.

Cost-effectiveness: an analysis undertaken alongside a clinical trial to determine the costs of producing a change in a particular health outcome. When health outcomes are expressed in terms of costs for increased quality of life (typically “quality adjusted life years” or QALYs, see below), the analyses are actually “cost-utility analyses”.

#### Cross sectional study: This is a study design that involves surveying a population at a single point in time. It can be used for assessing the prevalence of utilisation of treatment or of a specific condition in the population.

Delphi study: This is a consensus process designing to elicit expert opinion by asking participants to provide quantitative ratings, for example of the criteria for judging whether an acupuncture intervention is adequate. There may be several rounds of ratings, with confidential feedback to participants between rounds of both their rating and the group average rating. When this process also involves a face-to face meeting, then it is sometimes called a nominal group technique. By giving equal weight to the views and ratings of each participant there is less risk that individuals or sub-groups will dominate. This type of consensus approach is frequently used to develop criteria or guidelines where the literature-based evidence is inadequate.

#### Ecological validity: this is the extent that the patients, practitioners and style of acupuncture within a study reflect the context of normal practice. This is similar to external validity, see below.

Effectiveness: a therapeutic effect when the intervention is delivered as it would be in practice in the real world. These trials are usually simpler in design than efficacy trials as they allow participants to accept or reject the intervention and may include flexible treatment protocols and broader inclusion criteria. These studies have high external validity. Effectiveness trials may or may not involve “sham” or “placebo” controls – the definition is not concerned with that point.

Effect size: The effect size is a measure of the strength of the relationship between two variables. It is the difference in outcomes between the experimental and control groups using the standard deviation as a denominator to provide a dimensionless measure of the effect of the intervention. This difference between the two means, divided by the standard deviation, is also known as the standardized mean difference. In practical situations, effect sizes are helpful for making decisions: 0.2 is a small effect, 0.5 is a moderate effect and 1.0 is a large effect.

Efficacy: this is the therapeutic effect in an ideal, highly controlled setting with optimal administration of treatment. Efficacy trials are designed to include participants willing to adhere to the treatment regimen so they can determine whether the treatment works among those who receive it. It is worth noting that the use of “sham” or “placebo” controls does not itself constitute an efficacy study.

Equipoise:this is an ethical principle in medical research, when randomly assigning patients to different treatment arms, that the patient, the researcher and the person delivering the treatment should have no reason to prefer any one arm over the other.

Explanatory trials: A trial undertaken to establish the mechanism by which a therapy such as acupuncture may work. It controls for non-specific elements in order to evaluate the specific effects associated with the “active ingredients” of a treatment. Explanatory trials of acupuncture typically evaluate the effects of needling “in the correct location” compared to some form of “sham” acupuncture. Typically, these trials have stringent inclusion criteria and “objective outcomes” if possible. These design characteristics lead to high internal validity. The term “explanatory trial” should not be confused with “efficacy trial”.

#### External validity (generalisabilty): This is a measure of the validity of the results of a trial beyond that trial. For example, a study is externally valid and generalisable if its results are applicable to people encountered in regular clinical practice. This depends on the population from which the trial population was drawn, not just the entry requirements for the trial. External validity is enhanced when the acupuncture provided within a trial is representative of the acupuncture that is normally provided in routine practice.

Forest plot: These are used in meta-analyses to show graphically the variation in outcome between the individual studies under review. The plot also provides an estimate of the overall effect based on combining the results of all the relevant studies.

#### Generalisabilty: see external validity above.

#### Internal validity: A study is internally valid if it is designed and carried out in a way that provides results that are unbiased and gives an accurate estimate of the effects associated with the intervention.

Heterogeneity: This term is used in the context of meta-analyses as a measure of the dissimilarity between studies. This dissimilarity can be due the use of different statistical methods (statistical heterogeneity), or as a result of differences in the patient populations, treatment characteristics or outcomes measures (clinical heterogeneity). Pooling of data in meta-analyses may be unreliable or inappropriate when the trials are too heterogeneous.

#### Intention to treat (ITT) analysis: This is the method used to analyse the data for all participants based on the group to which they were (randomly) allocated rather than based on the actual treatment they received.

Kappa: known as Cohen’s Kappa, this statistic is used when two (or more) people provide a rating on a scale. On the basis that some agreement will occur by chance, the Kappa is a measure of the agreement between raters beyond chance: a Kappa of 1 is 100% agreement while a Kappa of 0 means no agreement beyond chance.

Manualisation: the process by which the description of an intervention for a clinical trial is sufficiently accurately described in a protocol for it to be delivered appropriately and replicated if required.

Mechanism: the method by which a cause induces an effect. In the case of acupuncture, the process by which needling might induce changes in outcome. To be distinguished from correlation which identifies association rather than causation.

Meta-analysis: This a quantitative method to summarise the results of several clinical trials in a single estimate, weighted on the basis of the size or quality of each trial. This is different from a systematic review which is an explicitly systematic search and appraisal of the literature.

Model validity: the model validity or “fit” is a measure of the extent that the assumptions underlying the research design are congruent with the practice or intervention being tested. Good model validity will be reflected in research methods that adequately address the unique healing theories and therapeutic contexts.

Non-specific effects: in clinical trials, the non-specific effects will depend on the hypothesis being tested, which can vary from experiment to experiment. In acupuncture trials, the non-specific effects include those effects that result from all components of treatment that are generic to the clinical encounter, such as empathy and rapport, but exclude the effects due to components of the treatment process that are integral to the theoretical underpinnings of acupuncture. When acupuncture is considered as a more complex intervention, there might be potentially active components beyond needling that are specific to acupuncture and therefore would contribute to the specific effects (see below).

Pilot studies: a pilot study is any study that is specifically designed as a forerunner of a more definitive study, for example a randomised controlled trial (RCT), to help develop and design it. Often, pilot studies mainly address the feasibility of the RCT, and so sometimes pilot studies are given the generic name feasibility studies.

Placebo: In explanatory clinical trials, the “placebo” is often used to describe the substance that is given to the control group. Ideally the placebo is identical in appearance, taste and feel to the intervention. Moreover it must not have any of the specific (active) effects that are associated with the experimental treatment. Placebo treatments are often referred to as sham treatments when the intervention is not pharmacological. Placebos are not the same as giving no treatment, as they are expected to mirror all aspects of the experimental intervention except those that are specific (active). Therefore placebos can have a physiological impact. The use of a placebo arm in a clinical trial depends on the research question.

#### P value: When comparing the effects of two interventions, and assuming there is no real difference between the effects, then the p value is the probability that an observed difference occurred by chance. The result is conventionally regarded as being "statistically significant" if this probability is less than one chance in 20, i.e. a P value less than 0.05.

Pragmatic trials: Also known as practical trials or management trials, pragmatic trials are designed to measure the overall benefits of a routine treatment, providing results that are directly applicable to normal practice. Unlike the explanatory randomised controlled trials (to evaluate efficacy under ideal conditions), pragmatic trails do not aim to separate the treatment out into specific and non-specific components, and usually compare a treatment to another accepted treatment (i.e. an active control), for example to standardised care or to no treatment. Pragmatic RCTs aim to recruit a population that is representative of those who are normally treated, so these trials tend to employ broader inclusion criteria. Results by are analysed on an "intention to treat" basis (see above). Pragmatic trial designs are useful when conducting cost-effectiveness analyses (see above).

Prospective studies: These are studies based on a pre-defined set of data that are collected over a specific period of time.

Psychometrics: this refers to the methods used to develop and test out an outcome questionnaire to ensure that it performs well and that it measures what it is designed to measure.

Quality Adjusted Life Years (QALYs): The Quality Adjusted Life Year (QALY) combines the quantity and quality of life on the basis that one year of perfect health-life expectancy is worth 1, and anything less than perfect life expectancy is worth less than 1, i.e. between 1 and zero (equivalent to death). When benefits are gained from a treatment, QALYs can be used to quantify this in terms of quality of life for the patient. Although QALYs gained are a fairly crude measurement, they can be compared across different interventions and therefore can enable decisions to be made on the allocation of resources where choices need to be made.

Qualitative study: an investigation that provides data on qualities rather than quantities, for example an investigation into the experience of an acupuncture consultation, from the perspectives of patients or acupuncturists. Qualitative data can be collected to provide information on perceptions of patients, acceptability of treatment, factors associated with satisfaction, etc, all of which can complement results based on quantitative data.

Quantitative study: an investigation that provides data on quantities rather than qualities. Quantitative data lends itself to statistical analyses.

Quasi-randomised controlled trials: These are trials where the method of allocating participants to different treatements is not truly random; for example, allocation by day of the week, by date of birth, or by alternating the order in which participants are included in the study. The limitation of this method is that the allocation is not adequately concealed compared with randomised controlled trials that include adequate allocation concealment (see randomised controlled trials below).

Randomised controlled trial (RCT): this is a method for randomly dividing patients into two (or more) equivalent groups, and testing for differences in outcome between a group receiving an intervention and a group receiving a control. This design provides for an assessment of the relative effects of interventions based on whether there are statistical differences in outcomes. The purpose of the randomisation and control are to minimise bias in order to better ascribe any differences in outcome to the intervention. The allocation to groups should be concealed from those providing treatment, so as to minimise potential interference with the random nature of the allocation process. Randomisation is likely to have been successful if the different treatment groups have the same characteristics at baseline, i.e there should be the same number of men and women, or older or younger people, or different degrees of disease severity. For descriptions of explanatory and pragmatic randomised controlled trials, see above.

Regression: this is the process by which any group of patients will tend to regress (return) to more “normal” levels over time. This is because patients often consult at a time when their symptoms are more severe, and the natural history of the condition may result in recovery over time anyway, regardless of the intervention. Sometimes such changes over time observed in a placebo group are called ‘non-specific’ or placebo effects. Regression to the mean is a statistical artefact whereby extreme values when measured again later tend to regress towards he mean. In theory the effects of regression are factored out in randomised controlled trials.

Reliability: this is a measure of the consistency of measurement, which could be consistency between different measures, or consistency of the same measure when repeated over time.

Responder ratio (RR): The ratio of the response rate among those receiving an intervention to the response rate of those in the control group. In acupuncture trials for example, if more patients respond to acupuncture than to the control, then the responder ratio is greater than 1, however if more patients respond to the control then the responder ratio is smaller than 1.

Retrospective studies: These are studies based on previously collected data.

Sham acupuncture treatment: An ideal sham acupuncture treatment will be identical in appearance and experience to acupuncture but lack any of the treatment specific effects. Various types of sham acupuncture have been used as controls: superficial needling at the same acupuncture points, superficial or deep needling at non-acupuncture points, skin tapping with cocktail sticks, stage dagger needles (that retract inside the handle of the needle), etc. The adequacy of the sham intervention is often unclear as there is no accepted delineation of what precisely are the acupuncture specific effects that one does not want to include in the sham intervention.

Significant effects: Significant effects are said to occur when there is a significance level of less than 5% (or P<0.05), i.e. when an observed difference between interventions would only occur by chance one time in 20. This is considered identical to having a 95% confidence interval that does not include the value corresponding to no effect. The convention is that when the word “significant” is being used, it is in this statistical sense.

Specific effect: in clinical trials, the ‘specific’ effect under examination will depend on the hypothesis being tested, which can vary from experiment to experiment. In acupuncture trials, the specific effects are often assumed to be the therapeutically active effects resulting from the insertion of acupuncture needles at precise points. When acupuncture is considered as a more complex intervention, there might be potentially active components beyond needling that are specific to acupuncture and therefore would contribute to the specific effects.

Standardized mean difference: the same as an effect size (see above).

Stratified randomisation: Where there is a known characteristic of patients that is likely to be a predictor of outcome, then stratified randomisation can be used to ensure that equal numbers of participants with this characteristic are allocated to each comparison group. Stratified randomisation is usually performed by using a separate randomisation for those with and those without the characteristic.

#### Systematic review: This is a review with specified and appropriate methods to address a defined research question by systematically identifying, appraising, and summarising studies, usually randomised controlled trials.