Attribution A judgement on whether one event was caused by another. Usually whether an *impact* was caused by an *intervention*. Will depend on whether other explanations can be eliminated and whether the *intervention* can credibly be seen as the cause.

Attrition The degree to which a study fails to include all the intended subjects due to factors such as drop-out or inability to contact them. May threaten the comparability of *treatment* and *control groups* and how far these remain representative of the *target group*.

Audit A quality assurance process that checks actions and procedures against guidelines and standards.

Blinding See *double-blind*.

Comparison group See *control group*.

Control group A group of people ('controls'), households, communities or other *units of analysis* who do not participate in the *intervention* being evaluated. Instead, they usually receive an alternative *intervention* (in which case the term *comparison group* may be preferable) or no *intervention* at all. Observations made on the controls are used to decide whether the *intervention* had an *impact* on the *treatment group*(s).

Cost-effectiveness One *intervention* is more cost-effective than another if it achieves more of a desired *objective* for a given expenditure.

Cost-benefit In a cost-benefit analysis both the costs and the benefits of *interventions* are expressed in monetary terms. This enables us to assess whether an *intervention* gained more than it cost and whether an alternative *intervention* achieved greater benefits for each £ spent.

Double-blind Research designs in which neither the subjects nor those taking measures from them know which *intervention* (if any) the subject received. Eliminates bias due to expectations or preconceived views. For the same reason, researchers may also be 'blinded' to other variables, such as characteristics thought to make subjects more or less receptive to *interventions*. See *placebo*.

Drop-out See *attrition*.

Effectiveness The degree to which an *intervention* produces the desired *objectives* under everyday conditions typical of those in which it will usually be applied. Contrast with *efficacy*.

Efficacy The degree to which an *intervention* produces a desired *objective* under relatively optimal or ideal conditions. A measure of its potential benefits rather than what we can expect from it in normal conditions. Contrast with *effectiveness*.

Evaluation A systematic assessment of whether and/or how the *aims* and *objectives* of an *intervention* have been acheived. May also assess unintended *outcomes* or other *impacts*.

Experimental group See *treatment group*.

External validity The degree to which what is evaluated in a study (and the conditions under which it is evaluated) permit us to assume that similar *impacts* will be observed in everyday practice. Can be maximised either by limiting the claims made for the study's *generalisability* or by employing more *naturalistic* research designs. Contrast with *internal validity*.

Generalisability How far an *evaluation*'s findings will be replicated in similar situations. Normally the main issue is whether the results will apply outside the research context to everyday conditions.

Hypothesis A formal prediction about what will happen as a result of an *intervention*. Such predictions are tested by the *evaluation*.

Impacts All the consequences of an *intervention* including intended and unintended *outcomes for* the *target group*.

Inputs The resources used to deliver an *intervention*, whether human, financial or physical.

Instrument An organised method for consistently collecting information such as questionnaires, guidelines for interviews and making observations, and protocols for testing urine and saliva. Because evaluations depend critically on how well they measure *outcomes* and other variables, instruments should be *objective*, *reliable* and *valid*.

Internal validity The extent to which the research design enables us to decide whether the *intervention* caused the observed *impacts*. The controls needed to achieve high internal validity often distance a study from real-world conditions, threatening its *external validity*. Internally valid studies are usually best suited to demonstrating *efficacy*. Contrast with *external validity*.

Intervention A policy, programme, service or project designed to bring about specified change to *target* areas or *groups*.

Longitudinal Research designs which aim to assess and reassess the same subjects at several time periods. For *evaluations*, the benefit of such designs is that they permit changes in each subject to be assessed against earlier measures taken from the same subject. See *prospective*.

Mediating variables Variables affected by the *intervention* which help cause the *outcomes*. For example, ability to refuse drug offers is increased by some prevention programmes and in turn is thought to lead to reduced drug use. When *outcomes* are hard to measure, changes in mediating variables may be used as a proxy for assessing the *intervention*.

Meta-analysis A study which uses recognised procedures to amalgamate results from several studies of the same or similar *interventions* to arrive at composite *outcome* scores.

Milestones Key stages in the *intervention* process which underpin later *outcomes* and which can be documented and monitored. For example, numbers attending for assessment or retained for a set period or the proportion of the target group reached by a campaign.

Monitoring An ongoing process involving the continuous or regular collection of key information about an *intervention*'s *inputs*, *outputs* and *outcomes*. This data may feed into a broader *evaluation*.

Naturalistic Describes a study of an *intervention* in 'real-world' conditions with minimal research interference, eg, without specially selecting subjects or controlling the quality of the *intervention*. Most appropriate to *effectiveness* trials. Often the only feasible approach in the light of resource constraints and ethical considerations which preclude allocating subjects to potentially inappropriate *interventions* or to none at all.

Null hypothesis The assumption tested by *statistical* procedures that a set of observations occurred purely by chance. In the current context, the null hypothesis usually amounts to the assertion that an *intervention* produced no *outcomes* or that there was no difference in the *outcomes* produced by two or more *interventions*.

Objectives Intended *outcomes* of an *intervention* which indicate that it has acheived its *aim*. Ideally specific, measurable, and attached to a timescale, in which case they can be expressed as targets.

Objectivity With respect to an *instrument*, the degree to which different people applying or scoring it in the same circumstances on the same subjects would register similar values. An aspect of *reliability*.

Odds ratio An odds ratio of 1 (the break-even point) suggests that the intervention is no better and no worse than doing nothing, below 1 that it is worse, above 1 that it is better.

Outcome evaluation An *evaluation* (or the element of an *evaluation*) which systematically records the *outcomes* of an *intervention*. Colloquially, whether the *intervention* 'worked'. Contrast with *process evaluation*.

Outcomes Intended or unintended end product of the *intervention* in the *target group*, eg, changes in substance use, infection control, reduced crime. If these match the *objectives* the *intervention* has worked.

Outputs Records or indicators of the level of throughput or activity of a service such as counselling sessions provided, level of occupancy of a residential service, training sessions provided. To be distinguished from *outcomes*.

Placebo A dummy *intervention* which mimics but lacks the presumed active ingredient of the *intervention*. Used to prevent subjects' expectations or preconceptions of the *intervention* systematically biasing *outcomes*. It is often impossible to construct a placebo condition when testing psychosocial *interventions*. See *double blind*.

Process evaluation An *evaluation* (or the element of an *evaluation*) which systematically documents the planning, implementation and delivery of an *intervention*. This may be as part of an attempt to establish its practicality (a feasibility study) or to elucidate how and why any observed *impacts* may have occurred. Colloquially, *how* the *intervention* 'worked' or why it did not. Contrast with *outcome evaluation*.

Prospective A study in which the subjects are recruited (and normally baseline measures taken) before the *intervention* takes place. Advantages usually include enabling *attrition* to be accounted for and *impacts* to be assessed by comparing measures taken after the *intervention* with those taken before.

Randomised controlled trial A study in which subjects are allocated at random to different *interventions* and/or to *intervention* and *control* groups. The intention is to eliminate the possibility that any *impacts* arose due to differences between the subjects in these groups rather than the *intervention*. Such studies are rare and (since self-selection or referral to *interventions* are the rule in practice settings) may suffer from low *external validity*.

Reliability A highly reliable *instrument* will deliver near identical results when applied repeatedly to the same subjects under the same conditions, and will do so even when different people administer and score the test. An *instrument* is unreliable to the degree to which measures taken with it may vary even when what it is supposed to be measuring has stayed the same.

Sensitivity In relation to a test, the proportion of people with the condition being tested for who are correctly identified. An aspect of *validity*. Contrast with *specificity*.

Specificity In relation to a test, the proportion of people *without* the condition being tested for who correctly test negative. An aspect of *validity*. Contrast with *sensitivity*.

Spontaneous remission Also termed 'regression to the mean'. The tendency for extreme or unusual behaviour (or attitudes, etc) to revert to more usual levels without formal *intervention*. Particularly relevant to therapeutic *interventions* as people often seek help when their problems have become unusually severe.

Statistical significance The findings of a study are accepted as statistically significant when they are very unlikely to have occurred by chance. The cut-off point is normally set at less then 1 in 20, expressed as a probability of less than 0.05 or '$p<0.05$'. If lower probabilities emerge we assume that something other than chance caused the results.

Statistical tests Accepted arithmetical methods to determine the probability that a set of observations occurred by chance. When this probability is below a certain level the observations are accepted as *statistically significant*. Such tests are important as unexpected causes of variation in *outcomes* could lead to unjustified conclusions about how well an *intervention* worked.

Target group The people, households, organisations, communities or other identifiable entities which an *intervention* is intended to affect. The degree to which the changes occur in this group constitute the *outcomes* of the *intervention*.

Treatment group People, households, organisations, communities or any other identifiable entities which receive an *intervention* as opposed to the *control group*. The term 'treatment' does not imply a medical or therapeutic *intervention* and may be replaced by 'experimental' or '*intervention*'. Contrast with *control group*.

Unit of analysis What constitutes a 'case' or 'subject' in the study. Usually an individual, but may be a group, a service, a family, a class or a school. To avoid mistaken conclusions, units *randomised* to *treatment* and *control* groups should correspond to those used to measure *outcomes*.

Validity With respect to an *instrument*, the degree to which it measures or otherwise reflects what it is supposed to measure. With respect to an *evaluation*, the degree to which conclusions drawn from the data correspond to reality; see *internal validity*, *external validity*.

Technical terms relating to evaluation

Standard definitions may have been adapted to fit the context of evaluations of interventions in the drug and alcohol fields. Terms defined elsewhere are italicised.