GLOSSARY

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.

Attribution Attributing *impacts* to causes. Usually whether a *statistically significant* result was caused by the *evaluated* intervention. The degree of confidence in this attribution will depend on whether alternative explanations can credibly be eliminated and whether the intervention can credibly be seen as the cause.

Attrition The number of subjects recruited into the study who did not receive the intended intervention or were not assessed. Can occur at various stages from recruitment to follow up and may threaten the continued comparability of *treatment* and *control groups* and otherwise weaken the *internal validity* of the study.

Blinding NEW 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 receive either no intervention or none relevant to the *outcomes* being assessed, or an alternative intervention (for the latter the term *comparison group* may be preferable). Observations made on the controls form the baseline against which changes in the *treatment group*(s) are assessed to determine whether the intervention had an *impact* and whether this is *statistically significant*.

Cost effectiveness The more cost-effective an intervention is the greater are the desired *outcomes* for a given expenditure on intervention.

Double-blind NEW Research designs in which neither the subjects nor those involved in taking measures from them know which intervention (if any) the subject has received, eliminating bias due to expectations or preconceived views. Blinding can also be applied to other variables knowledge of which might result in bias, 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 *outcomes*. When contrasted with *efficacy*, the extent to which it does so under conditions typical of those in which it will usually be applied.

Efficacy The degree to which an intervention produces the desired *outcomes* under relatively optimal or ideal conditions such as with expert, well trained staff, and selected subjects. Contrast with *effectiveness*.

Evaluation One formal definition is: "The systematic application of social research procedures in assessing the conceptualisation and design, implementation and utility of social intervention programmes." Less formally, the systematic attempt to assess an intervention in terms either of its feasibility or whether or how it contributes to desired *outcomes* or other *impacts*.

Experimental group See *treatment group*.

External validity UPDATED The degree to which what is evaluated (and the conditions under which it is evaluated) in a study permit us to assume that similar *impacts* will be observed when the intervention is applied as intended. Normally the extent to which research findings can be extrapolated to everyday non-research contexts. 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 NEW How far an evaluation's findings will be replicated in similar situations not actually studied. Normally the main issue is whether the results will apply outside the research context to everyday conditions, in which case *naturalistic* designs may be appropriate.

Hypothesis The predicted *outcome* of the intervention, normally based on theory, unstructured observations or previous research.

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

Instrument A structured method for collecting information such as questionnaires, interview and observation schedules and biometric tests of urine and saliva. In qualitative research, instruments should be *objective*, *reliable* and *valid*. For its results to be subject to *statistical testing* the instrument should produce a numerical score or a means of ranking or categorising the phenomenon it purports to record.

Internal validity The extent to which the research design enables conclusions to be drawn about whether the intervention caused the observed *impacts*. The higher this is, the more adequately can the study test the *hypothesis*. Depends on there being no relevant differences between *treatment* and *control* conditions other than the intervention. Achieving this may adversely affect *external validity*. Internally valid studies are usually best suited to demonstrating *efficacy*. Contrast with *external validity*.

Longitudinal NEW Research designs which aim to assess and reassess the same subjects at several time periods. In an evaluation context the benefit of such designs is that normally (by linking measures to subject identifiers) they permit changes in each subject to be assessed against earlier measures taken from the same subject. See *prospective*.

Mediating (or intermediate) variables Variables lying between the intervention and the anticipated *outcomes* in a hypothesised causal chain. With respect to drug and alcohol use, some examples are intention to use drugs (prevention), treatment retention and therapeutic alliance (treatment), and intoxication (community safety).

Meta-analysis A study which uses recognised procedures to amalgamate quantitative results from several studies of the same or similar interventions to arrive at composite *outcome* scores. Usually undertaken to enable the intervention's *effectiveness* to be assessed with greater confidence than it could have been on the basis of each individual study.

Milestones Key stages in the intervention process which underpin later *outcomes* and which can be documented and monitored. In treatment, may be numbers attending for assessment, retained for a minimum period, or engaging in aftercare. In prevention similar stages may be identified such as proportion of the target group reached, retaining awareness of the intervention's message, or engaging in recommended activities.

Naturalistic NEW Study of an intervention in 'real-world' conditions, eg, without randomising subjects and allowing the intervention to occur as it would outside the research context. Such studies typically observe and measure what happens normally rather than manipulating inputs in order to link these to *outcomes*. 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.

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*.

Outcome evaluation An *evaluation* (or the element of an *evaluation*) which systematically records whether and to what degree the intended *outcomes* of the intervention were achieved. Colloquially, whether the intervention 'works'. Contrast with *process evaluation*.

Outcomes The intended end product of the intervention or service, eg, changes in substance use or problems, infection control, reduced crime. To be distinguished from changes in *mediating variables* and *outputs*.

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 and attended. To be distinguished from changes in *mediating variables* and *outcomes*.

Placebo NEW 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. Such observations can be important in *attributing* both *outcomes* and *impacts*. Colloquially, *how* the intervention 'works' or why it did not. Contrast with *outcome evaluation*.

Prospective NEW A study in which the subjects are recruited (and normally baseline measures taken) before the intervention takes place. Advantages over retrospective designs (when measures are taken only after an intervention) usually include enabling *attrition* to be accounted for and the *impacts* to be assessed through a before and after comparison.

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 systematic differences between subjects receiving and not receiving the intervention(s). Such studies are rare and may suffer from low *external validity* as self-selection or referral to interventions is more usual in natural settings.

Reliability A highly reliable *instrument* will deliver near identical results in repeated data collections with the same subjects tested under the same conditions, even when (see *objectivity*) the people administering and/or scoring the test are different. An *instrument* is unreliable to the degree to which measures taken with it may vary even when the phenomenon being measured has not changed.

Spontaneous remission Also termed 'regression to the mean'. The tendency for relatively 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 A set of observations is accepted as statistically significant when it is highly unlikely to have occurred by chance. The cut-off point is set by convention, normally at less then 1 in 20, expressed as a probability of less than 0.05 or '$p<0.05$'. If lower probabilities emerge from a well-designed study it is acceptable to conclude that something other than chance caused the results, ie, to reject the *null hypothesis*. However, there remains the issue of *attribution* – establishing what the 'something' was.

Statistical tests Accepted arithmetical methods to determine the probability that a set of observations (measures, scores, categories, ranks) occurred by chance. When this probability is below a certain level the observations are accepted as *statistically significant*. Such tests are important in *outcome evaluations* as extraneous causes of variation in *outcomes* could lead to unjustified conclusions about how well the the 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 intended changes occur in this group constitute the *outcomes* of the intervention. However, *impacts* may also be seen in non-targeted groups.

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. Often an individual, but may be a group, a service, a family, a class or a school. To avoid mistaken *statistical* conclusions, the unit used in *randomising* to *treatment* and *control* groups should correspond to the unit used to measure *outcomes*.

Validity With respect to an *instrument*, the degree to which it measures or otherwise reflects the phenomenon it purports to record. For example, whether the results of a questionnaire intended to measure recent drug use correspond to accepted or more direct indicators of the same phenomenon, such as a pre-validated *instrument* or urinalysis results. With respect to an *evaluation*, the degree to which conclusions drawn from the data correspond to reality. See *internal validity* and *external validity*.

Published by
The FINDINGS Partnership,
c/o Standing Conference on Drug Abuse, 32–36 Loman Street, London SE1 0EE.
Layout and design **Mike Ashton**. Printed by **RAP**, 201 Spotland Road, Rochdale OL12 7AF. Tel 01706 644981. Fax 01706 345996