

Fonagy, P. (2018b). Methodological consideration in evaluating the outcome of psychoanalysis. In M. Leuzinger-Bohleber & H. Kächele (Eds.), *An Open Door Review of Outcome and Process Studies in Psychoanalysis* (pp. 67-83). London: International Psychoanalytic Association.

Methodological considerations in evaluating the outcome of psychoanalysis

by Peter Fonagy

The justification of effectiveness studies in psychoanalysis

In this section we shall consider the current climate in health care services which is largely responsible for the drive for effectiveness research and briefly overview some of the methodological issues that confront these studies. In the last part of this section we shall overview studies of psychoanalytically orientated psychotherapies.

Evidence based medicine and its justifications

Reasons behind the insistence on evidence

Psychoanalysis is a clinical intervention. Its aims and ambitions, at least from the point of view of most patients, are clearly associated with those of other healing arts such as surgery, physiotherapy and osteopathy. Admittedly, this is just one aspect of the psychoanalytic enterprise, but one that is crucial to its standing within most of the cultures where it is practised. Over the last ten years, all aspects of medicine have come under scrutiny, where increasingly both commissioners and funders of medical intervention, as well as those managing and directing clinical services, have embraced the values of “evidence based medicine” (Sackett [et al.](#), ~~Rosenberg, Gray, Haynes, & Richardson~~, 1996). Clinical judgement is no longer accepted as sufficient grounds for offering medical treatment. Recommendations at national policy as well as at local health care provider level are expected to be based upon evidence of effectiveness. What factors account for this change?

Ostensible reasons

Evidence based medicine is founded on an ideal – that decisions about the care of individual patients should involve the “conscientious, explicit and judicious use of current best evidence”. Much is claimed in favour of this approach, particularly in North America and Western Europe. The arguments in favour of it include (a) the more effective use of resources, (b) improvements in clinician’s knowledge, and (c) better communication with patients (Bastian, 1994). From an ethical point of view, the strongest argument in support of evidence based medicine is that (d) it allows the best evaluated methods of health care to be identified and enables patients and doctors to make better informed decisions (Guyatt [et al.](#), ~~Sackett, Cook, & the Evidence Based Medicine Working Group~~, 1994; Hope, 1995). All these are good reasons but all were as relevant to medicine in the past as at the moment. So why the current emphasis?

The political background

The real driving force behind evidence based medicine is unlikely to be a genuine concern for the quality of care. The movement appears to be largely driven by financial consideration and the hope of health care organisation to be able to reduce escalating costs by focussing on the most cost effective option given a range of treatments. Governments and health funds find the notion of allocating health resources on the bases of evidence quite attractive. In North America, D.K. Eddy in an important editorial suggested that healthcare funds should be required to cover interventions only if there was sufficient evidence that they can be expected to produce their intended effects (Eddy, 1996). The Australian Health Minister, Dr Michael Wooldridge, adopted a very similar position stating “[we will] pay only for those operations, drugs and treatments that, according to available evidence, are proved to work” (Downey, 1997).

While we believe that evidence for psychoanalytic interventions are important to derive, we are sceptical about the pressures brought on psychoanalytic clinicians as it seems to us unlikely that even in the face of overwhelming evidence as to the benefits of this relatively expensive treatment, the resources would be available to provide psychoanalysis for a significant proportion of those who require it. We shall consider the specific issue of cost effectiveness separately. In this context it is important to review the philosophical basis of the search for evidence for psychoanalysis in order to gain perspective on the entire enterprise of outcomes research. Perron’s (2001) critique has covered some of these issues from a more general epistemological standpoint; here some additional conceptual and practical concerns will be briefly explored.

Philosophical concerns

Evidence based medicine represents a practical example of “consequentialism”. Consequentialism refers to the proposition that the worth of an action may be assessed by the measurement of its consequences. There are at least three problems with the consequentialist argument, all of which apply to psychoanalytic outcome research: (a) the difficulty in measuring outcomes, (b) the ownership of outcomes (whose interest should be considered?), (c) consequentialism may lead to unethical conclusions. We shall take these in turn.

Philosophical questions concerning the measurement of outcome

The first concern is with the measurement of outcome. It is indisputable that many important outcomes of any medical treatment are unmeasurable. Evidence based medicine claims to provide a simple logical process for reasoning and decision making: (a) systematic scrutiny of the available evidence, (b) drawing appropriate conclusions leading to (c) a clinical decision as to the appropriateness of a treatment. Within this framework, for any decision to be balanced, all relevant consequences of a treatment must be considered. Unfortunately, in the current state of methods of psychological measurement, many important outcomes can only be very inadequately measured. Psychoanalysis concerns complex internal states such as the degree of distress or pain experienced by an individual. Often these complex states are reduced to simpler, easily measurable ones such as depression (Beck *et al.*, Ward, Mendelson, Mock, & Erbaugh, 1961), anxiety (Spielberger *et al.*, Gorsuch, & Lushene, 1970) or total symptomatology (Derogatis, 1993). A valid objection to such measures (if used without sophistication) is that they are reified and researchers may conflate the measure with the phenomena they were aimed at quantifying. Thus, the BDI score is not depression and the total symptom distress score of the SCL-90 is not equivalent to mental pain. By having these

measurements we have not at all done justice to the complex cognitive, affective and physiological processes which are implicated by these terms.

Even if better measures were found for some of the domains of outcomes entailed in psychoanalytic treatment, other aspects of the process, such as an ethical life, a sense of purpose or social justice, may be inherently unmeasurable. Even more troublesome are key domains which are not even well defined, let alone measurable. One such is the “quality of life”. Attempts have been made to provide a metric for this, yet in the absence of a consensus as to what a reasonable quality of life might entail, it is hard to imagine how measurement is possible.

The philosopher Bernard Williams (1972) noted that values that can be quantified in economic terms, may require comparison with values which are not quantifiable. His comments may be easily extrapolated to the current situation of psychoanalysis in some countries: “Again and again defenders of such values are faced with the dilemma of either refusing to quantify the value in question, in which case it disappears from the sum altogether, or else of trying to attach some quantity to it, in which case they misrepresent what they are about and also usually lose the argument, since the quantified value is not enough to tip the scale” (p 103). Some outcomes of psychoanalysis may indeed be costed, but these may be some of the least important. The cost saved may not “tip the balance” in favour of psychoanalysis.

The ownership of outcome

The second common criticism concerns the ownership of outcome: “Whose outcome is the outcome of psychoanalysis, anyway?”. It may be in principle impossible to decide between the competing claims of different individuals. For example, a treatment that enhances the quality of life of one person may be deleterious to a spouse or an employer. This is particularly evident in the case of the psychoanalytic treatment of children where the treated child’s desired outcome may be in conflict with that of the parent’s, or indeed that of the sibling. Ideally, notwithstanding the insurmountable practical problems, all individuals significantly concerned with an analysand should be assessed as part of the outcome study. The research enterprise itself is clinician led. It is the clinician-researcher that decides whose outcome will form the basis of evidence based practice. Thus all outcome investigations, perhaps particularly that of psychoanalysis, will be arbitrary, and limited by the selection of the individual(s) on whom outcome is measured.

An extension of the arbitrariness problem of outcome ownership concerns the status of client choice as an indication of outcome. It could be argued that the client is in a privileged position relative to the investigator in determining whether the treatment is helpful. Interestingly, when user groups are asked they tend to strongly favour approaches to most mental health problems which are psychologically rather than pharmacologically based, or at least they plead for a greater emphasis on psychological help. When individuals perceive their difficulties arising out of psychosocial causes, they understandably seek redress in the same domain i.e. the interpersonal. It is also worth noting that psychoanalytic therapy often has greater *prima facie* acceptability than exposure-based cognitive behaviour therapy (for example with patients with OCD, Apter [et al.](#), Bernhout, & Tyano, 1984). Yet the desire of the user, “client satisfaction” is not generally acceptable as an adequate criterion for outcome. By this criterion, many treatments known to be ineffective and even harmful, (e.g. recreational drugs such as nicotine counteract anxiety) could be selected.

Psychotherapy researchers are particularly conscious of the danger of imposing ethnically rooted cultural biases on what is designated as “needing treatment” and to be a “good outcome” (Bernal [et al.](#), Bonilla, & Bellido, 1995). For instance, the achievement of selfhood

through the separation-individuation process is one of the cornerstones of psychotherapeutic interventions. Yet is Lasch (1978) correct that the emphasis on individual achievement in Western culture is excessive and that an appropriate submission to the goals of the family and community (Kagan, 1984) may be a far better indicator of healthy adaptation? Such differences are particularly acute in the area of child development and parenting. Rogler (1989) outlined some of the practical steps which culturally sensitive outcome research requires. In particular, it is important to ensure that interventions are consonant with the subjective culture of the ethnic group to which it is applied and that instruments used are able to integrate cultural meanings with the pertinent scientific categories. In reality, this is an ideal to strive for, but it is rarely achieved.

Ethical concerns

Finally, it is commonly asserted that a uniquely evidence based treatment approach can lead to activities which are at odds with common morality. A good example of this is the success of aversive conditioning and other punishment based techniques in behavioural control of individuals with “challenging behaviour”. The fact that there is evidence supporting the efficiency of these techniques cannot and does not make them right.

More generally, ethical concerns arise out of the implementation of randomised control trials. While such trials have the potential to prevent the propagation of worthless treatments, for example insulin coma therapy, they raise major ethical issues in the context of subject selection, consent, randomisation and the continuing care of subjects once trials are complete. Randomised control trials require the clinician to act simultaneously as physician and research scientist. Patients are simultaneously invalids and research subjects. It is questionable if the physicians’ moral responsibilities towards patients can be consistent with the recommendation that the patient should participate in a randomised control trial, principally because of this conflict of interest (Hellman & Hellman, 1991). It has been suggested that such trials may be recommended by the physician if clinicians are in a state of “therapeutic equipoise”, that is they are genuinely in doubt about the value of different interventions (Lilford & Jackson, 1995). Such equipoise may be achieved in the case of treatments with moderate affects which might otherwise be obscured by bias and random effects. However, equipoise may not be achievable when interventions have great benefits and risks and then alternative clinical procedures to be investigated by other methods.

Is therapeutic equipoise applicable to the recommendation of psychoanalytic treatment? Interestingly, neither psychoanalysts nor the opponents of psychoanalytic treatment believe that this is the case. Psychoanalytic clinicians are so firmly convinced of the appropriateness of 4 or 5 times a week treatment that they tend to consider it unethical to recommend less intensive alternatives. Sceptics, on the other hand, feel that the sacrifice demanded of the patient and his/her family is such that randomisation to a psychoanalytic arm is normally ethically unacceptable. In principle, the existence of these opposing views might somehow be combined to construct an attitude of therapeutic equipoise, but in reality it is simply tantamount to what may be an insurmountable obstacle facing a randomised controlled trial of psychoanalysis.

The status of concerns about evidence based medicine

Many other concerns could be raised about the appropriateness of subjecting psychoanalysis to outcome evaluation. We raise some concerns here in part to demonstrate our awareness of the issues and in part to underscore that the clamour for evidence should be met with caution and sophistication. It needs to be recognised that objections to research will not win the day. It is unlikely that the prevailing view which places controlled studies at the top of the hierarchy

of evidence will change no matter what the pressures of arguments. The complexities of issues surrounding resource allocation, the drive to seek certainty and simplicity at the level of policy making are such that alternative formulations will not be heard.

Psychoanalysis is not alone among medical treatments with a weak evidence base. Evidence to the standards required is available for relatively few medical interventions (Kerridge [et al., 1998](#)). The drive for an evidence base for the selection of treatment interventions will inevitably mean a biased allocation of resources to those treatments for which rigorous evidence of effectiveness is relatively easily collected or where funds are independently available to carry out more lengthy and complex effectiveness research. Brief therapy benefits from the former, pharmacotherapy from the latter. Psychoanalysis is further disadvantaged by the opposition to many of its fundamental propositions among fellow mental health professionals and influential leaders (Crews, 1995; Grünbaum, 1984; 1986; Webster, 1995). These kinds of considerations drive us to override our concern and accept the imperfect solution of outcome research with the overriding objective of preserving the discipline.

The best strategy available to us is to collect all the data available rather than enter an epistemological debate amongst ourselves. The debate is inaudible to those outside the discipline. Further, it would sap our energies when this is required for a collaborative effort to make the best case possible for psychoanalysis as a clinical method. Even those of us who are engaged in collecting evidence for the effectiveness of this discipline have major methodological as well as epistemological concerns. These should not be set aside, forgotten about, but nor should they become an alternative focus.

It should be remembered that the debate over the effectiveness of psychoanalysis is one of pragmatics not of principles. There is a clear danger that the therapy that is “without substantial evidence” will be thought by all to be “without substantial value” (Evidence Based Care Resource Group, 1994). Once this idea is allowed to flourish, a cultural change becomes inevitable, a change which at least temporarily has the power to stop the development of our discipline – through the rejection of psychoanalysis as the therapeutic choice, through discouraging young people from entering the profession and through bringing psychoanalytic contributions to mental health disciplines and other subjects into disrepute.

Methodological problems inherent to evaluation research

Research into psychoanalysis is inevitably a compromise between usual clinical procedures and the demands of scientific influence. Clear thinking about the applicability of research findings rests on an understanding of the nature of these compromises. In this section we shall briefly list some of the issues which must be taken into consideration in interpreting and evaluating evidence for the effectiveness of psychoanalysis. While these issues are well known and obvious to some, they may be less familiar to others. More important, we list them here in part to show that researchers are well aware of these problems and while not necessarily able to resolve the issues, at least it should be clear that they are working towards this end.

Efficacy versus effectiveness

The term efficacy refers to the results a treatment achieves in the setting of a research trial, while clinical effectiveness is the outcome of therapy in routine practice. The discrepancy arises because trials are required to show “internal validity” (Cooke & Campbell, 1979); that is, they permit causal inferences to be made on the basis of the observed relationship between the variables. In this context, the absence of a relationship must imply the absence of a cause.

Achieving internal validity normally requires modifications to clinical procedures, which are rarely seen in everyday practice. The most common of these are: (a) the selection of diagnostically homogenous patient groups, (b) the randomisation of these patients into treatments, (c) the employment of extensive monitoring of the patient's progress, (d) the careful specification of therapeutic procedures to be used and (e) the monitoring of their implementation. These requirements clearly pose a threat to "external validity", to the extent to which the inferred causal relationship between variables may be generalised. Thus demonstrations of efficacy are not necessarily demonstrations of effectiveness. The fact that a treatment is highly efficacious under strictly controlled conditions cannot be thought to mean that it will have the same value in the context of ordinary clinical practice.

This problem is by no means unique to the investigation of psychodynamic treatment. To take a simple example, a pharmacological agent with distinctly unpleasant but harmless side effects may be shown to have considerable efficacy in a double blind controlled trial. No one would be surprised that it proves to be ineffective in clinical practice since patients frequently and conveniently "forget" to take this pill. In the trial, serum levels were carefully monitored and subjects whose blood levels indicated that they did not take their drug were excluded from the analysis. The same applies in trials of psychological treatment. Frequently psychotherapy is not delivered in practice as well as it is in the context of a carefully monitored trial. By contrast trials may underestimate the effects of a therapy by randomly assigning patients to treatments they do not wish to have, whereas in clinical practice their preference would be carefully noted by their treating physician.

Spontaneous remission

As relatively few of the individuals who suffer from significant psychiatric morbidity have the benefit of any kind of professional help, it must be obvious that there are many roots to recovery which do not involve psychoanalysis, psychotherapy or indeed any kind of systematic intervention. What any treatment needs to demonstrate therefore, is that it is more effective than the natural processes of healing which human society provides (note for example Freud's famous comments about the therapeutic potential of Lourdes (Freud, 1933a). From a historical point of view, Hans-Jürgen Eysenck (1952) was the first to raise this issue in connection with psychoanalytic therapy. He claimed, on the basis of insurance statistics as well as Fenichel's Berlin I Study of the outcomes of the Berlin Psychoanalytic Institute, that more individuals recovered in a two year period when they were untreated than when they were treated in psychoanalysis. More recently, Forty years later, it was demonstrated that even using Eysenck's data a more sophisticated analysis reveals that whereas half of treated patients improved within a couple of months, only 2% of those untreated improved over the same time period (McNeilly & Howard, 1991).

Whatever the status of Eysenck's own figures, there is no doubt that spontaneous improvement rates are sizeable for most psychological disorders (Bergin, 1971; Lambert, 1976; Subotnik, 1975). For example, from naturalistic follow up studies we know that individuals with borderline personality disorder tend to "burn out" in middle age (Stone, 1990). Thus statements about the effectiveness of psychoanalysis cannot be made on the basis of clinical reports of individual cases, however successful – certainly not without unequivocal knowledge about the course of the disorder. Ideally the course of untreated individuals should be compared with those who receive treatment. It is impractical and unethical to withhold treatment from an individual for the duration of a longterm treatment such as psychoanalysis and this has posed major problems for those intending to carry out outcome studies. As psychoanalysis is not generally available it seems sensible to compare its effectiveness with either the best available alternative treatment or so-called "treatment as usual". The former

has the advantage of offering an apparently meaningful comparison from the point of view of a referrer or referring agency, but equally has the potential of prompting meaningless comparisons where the aims of treatment are not comparable and apples are being compared with oranges. Such comparisons also require that the researcher has comparable expertise with both the methods of treatment, as well as large sample sizes as the difference between the two methods is likely to be small. The alternative contrast with a treatment as usual group, has the advantage of telling us how much difference a treatment might make were it to be added to routine care but has the disadvantage of potentially great heterogeneity in the control group and inadequate information concerning the treatment received by the control group (Roth & Fonagy, 1996).

Strategies of psychotherapy research

The choice of a particular research methodology will always be a compromise, reflecting the intentions, interests (and resources) of investigators. Some of the major strategies used in psychoanalytic research, together with their strengths and weaknesses, will be considered in turn. A full account of these issues in psychotherapy research is given in Kazdin (1994).

Single case studies

The belief that knowledge based on groups of individuals is somehow more likely to be generalisable – that is, applicable beyond the specific locus of its discovery – than is the case for knowledge based upon individual cases, is fatally flawed (Fonagy & Moran, 1993). In single case designs the focus is on the individual patient rather than a group average, even where a group of patients were studied. Single-case studies may be descriptive or quantitative. The former group is well represented in the traditional psychoanalytic case history. The method has many strengths, including high communicative value, and the richness of description of particularly complex unconscious interactive processes between analyst and patient. There is no generally accepted format for these reports and the information included tends to be quite variable (e.g. Spence, 1994) which undermines generalisation. Attempts have been made to systematise such qualitative reports (e.g. Klumpner & Frank, 1991) but these have not met with general approval.

In comparison to descriptive accounts of single treatments, quantitative reports undoubtedly lack richness and depth but are more generally accepted because of the greater ease with which the reliability of the observation can be assessed. Within this latter group some are naturalistic reports of outcome or quasi-experiments (Cooke & Campbell, 1979), while others are reports of the experimental manipulation of interventions. In cases where appropriate baseline measures are taken, or where treatments are applied and withdrawn in a controlled manner, the patient acts as his/her own control. This methodology has been widely used by behavioural and cognitive-behavioural researchers (Morley, 1987; 1989), but is equally applicable to psychodynamic investigators (e.g. Fonagy & Moran, 1993) and to the investigation of process factors in therapy (e.g. Parry, 1986).

Single-case studies have a number of attractive features. They can be combined with the routine clinical practice of private practitioners, they do not (necessarily) require the research apparatus and personnel normally associated with group based research and can be conducted fairly quickly. While of great importance in the demonstration or refinement of clinical technique and especially in treatment innovation, the results of single case studies can be difficult to generalise to the broader clinical population (indeed the design is not intended for such a purpose). Patients are often highly selected (necessarily so where studies are aiming to show the effectiveness of a technique for particular clients). More fundamentally, however, interpretation of results is limited by the fact that (as will become evident in the body of this

report) therapeutic interventions have both general and specific impacts on the welfare of patients. A contrast intervention is required in order to be clear that any demonstrated benefits are attributable to specific therapeutic techniques – a strategy adopted in the randomised control trial.

Randomised Controlled Trials (RCTs)

In contrast to the single case study, RCTs explicitly ask questions about the comparative benefits of two or more treatments. Patients are randomly allocated to different treatment conditions, usually with some attempt to control for (or at least examine) factors such as demographic variables, symptom severity and levels of functioning. Attempts are made to implement therapies under conditions which reduce the influence of variables likely to influence outcome – for example by standardising factors such as therapist experience and ability, and the length of treatments. The design permits active treatments to be compared, or their effect contrasted with no treatment, a waiting list or a “placebo” intervention. Increasingly, studies also ensure that treatments are carried-out in conformity with their theoretical description – for example, ensuring that psychoanalytic treatments do not include cognitive-behavioural or supportive elements. To this end many treatments have been “manualised” (a process which specifies the techniques of the therapy programmatically), and therapist adherence to technique is monitored as part of the trial. There are obviously major problems in the manualisation of psychoanalytic treatment (Clarkin, 1998) but some progress has already been made on this front (e.g. Clarkin et al., 1999; Fonagy **et al.**, Edgecombe, Target, Moran, & Miller, 2000; Kernberg et al., 1989; Luborsky, 1984).

Though this design has the potential to distinguish the impact of treatments (and to provide a control for the effects of spontaneous remission), there are inherent limitations to this approach.

Problems of control groups

Although the ideal design of a treatment would be to contrast treatment to no-treatment, it is rarely the case that this is either ethically or practically possible. The alternative of offering a placebo treatment – one which is considered inactive, at least from the point of view of the active treatments offered – is beset by the difficulty of finding an activity which could be guaranteed to have no therapeutic element, which controls for the effect of attention and which is also viewed by patients as being as credible as a psychiatric intervention. Many recent studies restrict themselves to the comparison of active treatments; as evidence has accumulated for the general efficacy of therapy, institutional review boards (ethical committees) have become unwilling to sanction trials which could be seen to deprive patients of help (e.g. see Elkin, 1994).

Length of therapy

Setting up an RCT is a major undertaking, and consequently a great expense. Although there are exceptions, most trials limit the amount of intervention offered (frequently to around 16 weeks). While this may be appropriate for some therapies (principally behavioural or cognitive-behavioural approaches), psychodynamic therapists (e.g. Fonagy & Higgitt, 1989) could – and do – argue that the techniques they employ were never designed for delivery over such a short time-frame. Psychoanalysis is in most countries an open-ended treatment and it is hard to imagine forcing it into a frame where the number of sessions is determined independently of the individual treatment process.

Generalisability

Few RCTs achieve the implementation of psychological therapies under conditions which might be obtained in routine practice. As noted above, because they are characterised by a concern to maintain internal validity, their applicability could be seen as limited. For example:

- patients will have been selected to conform to diagnostically precise categories
- patients will have been exposed to multiple assessments
- therapies will be applied with some precision, often under supervision
- researchers will often be particularly enthusiastic and particularly expert in the techniques they employ.

Patient preference and random allocation to treatment

Patients are not passive recipients of treatment, and their preferences for differing forms of treatment may be critical to their participation in clinical trials (Brewin & Bradley, 1989). The bias introduced by consequent attrition from treatment is invisible within studies, but may be particularly relevant to clinical practice.

Open trials

This methodology is intermediate between the single-case design and the randomised control trial. Although entry to treatment may be governed by strict criteria, there is no control group. Such designs often reflect a more naturalistic treatment protocol than is the case with RCTs. At the simplest level such studies offer important information concerning:

- the likely benefit the average patient might derive from the treatment
- what features of presentation are likely to be associated with relatively good outcome
- how effective a particular service is in terms of outcome
- which aspects of a patient's problems are likely to be addressed by a treatment
- given a certain natural variability in treatment delivery, what aspects of treatment are associated with felicitous consequences and which are accompanied by equivocal outcomes.

Frequently two or more treatments for the same disorder, as practised in different settings, are contrasted. In principle, such a design could answer the question "what kind of patient benefits most from particular treatment protocols". In reality differences in case-mix and the failure to control specific components of treatment usually place drastic limitations on the implications which may be drawn from such studies. Given a sufficiently large data-set, it may be possible to derive conclusions about the relative value of treatments even in the absence of random assignment. However, studies on such a large scale are rarely possible.

Resolving conflicts between internal and external validity in research designs

We have already noted that a major problem for outcome studies of psychoanalysis is the tension between satisfying the demands of internal and external validity when developing research strategies. Designs have to reach a compromise between these factors; bridging the gap between them requires innovative attempts at integrating an apparent incompatibility between scientific rigour on the one hand and generalisability on the other. Single-case designs may come to play a more important role in this respect, since external validity is not an inherent problem in designs of this type (Kazdin, 1994). When replicated across randomly sampled cases, they have considerable generalisability. They can be employed to answer most

of the questions that concern researchers, such as the appropriateness of a particular form of treatment, the length of treatment required to achieve a good outcome, the relative impact of treatment on particular aspects of the problem or the relevance of particular components of treatment. However, there is one critical exception: within this research strategy patient and analyst factors are difficult to study. If there is no replication across subjects (patients and analysts), the design will not yield information about their influence on outcome.

Thus methodology which is truly adequate to the task of simultaneously assuring internal and external validity in psychoanalytic research has probably yet to be developed. In the meantime, the best – though possibly inadequate – answer lies in reviews (such as the present one), which include critical appraisal of likely threats to external validity posed by current research.

Other considerations

Follow-up

For most conditions the success of therapy may be measured by its ability both to improve patient functioning and to maintain that improvement after therapy ends. Although most trials report follow-up data, the length of follow-up can vary markedly between studies, sometimes being only a matter of weeks, sometimes years. The length of follow-up required to demonstrate a clinical effect is governed by the natural history of a disorder, which will suggest both the probability of relapse and the usual length of time between episodes. Therapeutic efficacy can only be demonstrated in the context of both factors and, for example, three month follow-up for a condition known to show greatest relapse over a period of one year would clearly be inadequate. This aspect of research design is particularly important for psychoanalytic investigations where so called “sleeper effects” have been frequently reported (e.g. Kolvin et al., 1981). The term refers to improvements observed after the termination of treatment. Termination is a complex time in psychoanalytic treatment with recurrence of the original complaints commonly reported.

Although this suggests that extended follow-up periods should be the norm, the longer a patient is followed-up the more difficult it is to ascribe change to their original treatment. In part this is because patients will might seek further treatment in the intervening period (e.g. Shea et al., 1992), and also because the relative impact of treatment in the context of life-experiences decreases over time. Ironically, the results of very prolonged follow-up, while desirable, may be difficult to interpret.

Finally, the stability of symptomatic change over the follow-up period may be an issue of concern in its own right. Monitoring of individual patients suggests that a proportion will change their symptom status more than once (e.g. Brown & Kulik, 1977; Shapiro et al., 1995). Reporting of group-averages tends to obscure this variability, leading to an over-estimation of longer-term outcomes in clinical practice.

Attrition

All clinical trials will lose patients at various points in treatment; the point at which they are lost will have differing impacts on validity. Early loss from a trial may disrupt the randomisation of treatment, threatening internal validity. Even where there is no differential attrition from treatments, it may be the case that significant attrition could lead to results being applicable only to a sub-group of persistent patients, threatening external validity. Alternatively, attrition rates across treatment conditions may not be random, and may reflect the acceptability of therapies, suggesting that attrition may be an important variable in its own right.

Significant levels of attrition will restrict the conclusions that can be drawn from a study, and complicate reporting of results. A number of statistical solutions to this problem are available to researchers which utilise the last available data-point to estimate the likely bias introduced by loss of patients (e.g. Flick, 1988; Little & Rubin, 1987). Alternatively data can be reported on the basis of an "intention-to-treat" sample, including all subjects entered into the trial, as well as presenting separate data for those completing all or a specified length of therapy (e.g. Elkin et al., 1989).

Meta-analysis

In the past 15-20 years, techniques have been developed to enable quantitative review of psychotherapy studies. Meta-analysis is a procedure which enables data from separate studies to be considered collectively through the calculation of an effect size from each investigation (Rosenthal, 1991).

Effect sizes are calculated according to the formula:

$$ES = \frac{M_1 - M_2}{S.D.}$$

S.D.

where

M_1 = the mean of the treatment group

M_2 = the mean of the control group

S.D. = the pooled standard deviation

The terms M_1 and M_2 can stand for the means of any two groups of interest, such as psychotherapy contrasted against a waiting list control, or equally could be the comparison of two forms of psychotherapy. Because this technique converts outcome measures to a common metric, individual effect-sizes can be pooled. In addition to examining the contribution of main effects such as therapy modality, effect-sizes for any variable of interest can be calculated, such as the impact of methodological quality or investigator allegiance on reported outcomes (e.g. Robinson et al., Berman, & Neimeyer, 1990; Smith et al., Glass, & Miller, 1980).

Effect sizes refer to group differences in standard deviation units on the normal distribution. Their intuitive meaning is made clearer by translating them into percentiles, indicating the degree to which the average treated client is better off than control patients. Thus an effect size of 1.0 corresponds to a result where 84% of the treated group are better off than the average control patient.

Meta-analysis is a powerful research tool, but some have been critical of the technique (e.g. Wilson & Rachman, 1983). Common criticisms include:

the fact that reviews do not include single-case studies

the inclusion of studies of questionable methodological adequacy

the inclusion of studies not directly relevant to clinical issues, such as analogue studies, and trials of patients whose symptoms are not clinically significant or of great severity

the fact that analyses can multiply sample measures taken from the same patient and from the same study leads to effect sizes computed on the basis of dependent data

the fact that using average Z scores assumes that outcome measures are appropriately measured on an interval scale, and that their distribution may be assumed to have insignificant skewness and kurtosis

sampling of studies will be biased by the tendency for editors and authors to favour positive results

not all meta-analyses weight the means for sample size.

A major difficulty is, however, that the effect size statistic can only speak to treatment effects for the average client, and though this is informative of general treatment effects, further elaboration of therapeutic impacts is usually required to detail the more specific effects of treatment.

Problems associated with the use of statistical tests in psychotherapy research

Clinical and statistical significance:

Much of this report is based on journal articles examining the truth of the null-hypothesis – in essence the proposition that psychoanalysis has no effect, or no effect greater than a control treatment. It is conventional to report the statistical significance of differences between treatments in terms of a confidence level of $p < 0.05$ or $< .01$. However, researchers may be able to reject the null-hypothesis at relatively high levels of statistical significance without simultaneously demonstrating that this finding is worthy of clinical attention (Kukla, 1989). Demonstration of statistical effects may not be equivalent to a clinically significant therapeutic change, and there are a number of strategies which have been used to detect this (discussed further in Kazdin, 1994):

Comparison of patient change with normative samples

Measurement of the extent of individual change by reference to a criterion measure of change; for example, that treated clients should be 2 standard deviations from the mean of the untreated group (Jacobson & Truax, 1991)

The use of a criterion of recovery which enables categorical rather than continuous scoring of outcomes; for example, considering all individuals scoring as low as 75% of the normal population to have benefited from the treatment (e.g. Elkin et al., 1989).

The clinical significance of change is central to the evaluation of psychotherapy outcomes; though recent investigations are more likely to report data in this form, such measures are not always available.

Multiple data sampling and type-I error

Researchers frequently report numerous results of statistical significance without being clear how each test relates to the prediction they are examining. Dar and colleagues (Dar, Serlin, & Omer, 1994) illustrate this problem by suggesting a hypothetical study in which two treatments for flying phobias are contrasted, with levels of anxiety and coping skills being the dependent variables. In practice there may be a number of procedures for measuring these variables, all of which are likely to be intercorrelated. Each of these variables could be examined separately, though in reality there are only two hypotheses under investigation – the impact of the treatment on anxiety and its effect on coping skills. More than two statistical analyses are therefore redundant, and represent an overstatement of the data available to the researchers. A real-life example of this process is the much-cited National Institute of Mental Health study of treatments for depression (Elkin, 1994) which shows statistical significance on only some of a relatively large family of variables pertaining to dysfunctional emotional states. A consequence of multiply-sampling related data-sets is to increase the risk of Type I errors –

rejecting the null-hypothesis when that hypothesis is false (in practice, for example, claiming that one treatment works better than another when in reality both work equally well).

Because it is well recognised that a series of measures tapping similar domains may be inter-related, investigators often employ multivariate tests, which permit some understanding of relationships between dependent measures. Though this procedure overcomes some of the problems noted above, problems can arise where multivariate tests which indicate overall significance are then followed by univariate tests. Not only does this increase the risk of Type I error, but results can be difficult to interpret, once again because of possible relationships among variables under test.

Atheoretical analysis

Dar et al. (1994), in a review of the use of statistical tests in psychotherapy research from the 1960s to the 1980s, note a high level of inappropriate significance testing, which they attribute to the pragmatic concerns of psychotherapy researchers. The determination to find statistically significant associations is seen by them as motivated by "a flight from theory into pragmatics". As psychotherapy research frequently has very little theoretical guidance leading to meaningful hypotheses and testable predictions, there has been an explosion of exploratory procedures, leading to a state of affairs where, even in the best journals, "much of the current use of statistical tests is flawed". Psychoanalytic outcomes research is sadly no exception to this trend and many of the studies included in this review have undoubtedly over-exploited their data.

Statistical power

Statistical power is the extent to which an investigation is able to detect differences between samples when such differences exist in the population – in other words when there is a true difference between the groups under test. Power is a function of:

the criterion for statistical significance, or alpha level

sample size

effect size, or the magnitude of the difference that exists between the groups.

Statistical power in perhaps the majority of trials of psychoanalysis may be relatively weak, primarily because of low sample sizes (Kazdin, 1994). Cohen (1962) distinguished three levels of effect size (small=0.25, medium=0.50 and large=1.0), and evaluated the ability of published studies to detect such differences at the conventional alpha level of $p < 0.05$. Power within these studies was generally low – for example, studies had a one in five chance of detecting small effect sizes, and less than a one in two chance of detecting medium effect sizes. Despite the cautionary note struck by Cohen's paper (1988), and the date of its publication, Dar and colleagues (1994) found that a significant proportion of even recent research continues to neglect these issues. Most particularly, there continues to be a neglect of measures of effect size in favour of citing statistical significance. The problems inherent in this procedure can be readily illustrated by considering a study with a large sample but a small effect size; although statistical significance may well be achieved this does not speak to the magnitude of the effect, nor its likely reliability or validity. In psychoanalytic studies the reverse scenario is often more likely: too few subjects being compared reducing the likelihood of the demonstration of significant changes, even when such changes are present.

It should be clear that all of the above issues threaten the external validity of psychoanalytic research. Dar et al. (1994) detail a number of strategies for ensuring that such threats are minimised; for example, by employing theory-guided predictions, planned rather than post-hoc

statistical decisions, reduced use of omnibus multivariate techniques, stricter control of type-I error rates by using single rather than multiple tests, employing “families” rather than a multiplicity of hypotheses, the avoidance of step-wise statistical procedures and testing of hypotheses not against a difference of zero but rather against a predetermined interval. While these suggestions are well taken, the opportunities for psychoanalytic research are at the moment so few that many of these methodological niceties will

References

- Apter, A., Bernhout, E., & Tyano, S. (1984). Severe obsessive compulsive disorder in adolescence: A report of eight cases. *Journal of Adolescence*, 7, 349-358.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (1st ed., pp. 217-270). New York: Wiley.
- Bernal, G., Bonilla, J., & Bellido, C. (1995). Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with hispanics. *Journal of Abnormal Child Psychology*, 23, 67-82.
- Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *British Medical Journal*, 299, 313-315.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5, 73-99.
- Clarkin, J. F. (1998). Intervention research: Development and manualization. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive Clinical Psychology* (Vol. 3, pp. 189-200). New York, NY: Pergamon.
- Clarkin, J. F., Kernberg, O. F., & Yeomans, F. E. (1999). *Transference-Focused Psychotherapy for Borderline Personality Disorder Patients*. New York: Guilford Press.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences*. Hillsdale: Lawrence Erlbaum.
- Cooke, T., & Campbell, D. (1979). *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Crews, F. (1995). *The Memory Wars: Freud's Legacy in Dispute*. London: Granta Books.
- Dar, R., Serlin, R. C., & Omer, H. (1994). Misuse of statistical tests in three decades of psychotherapy research. *Journal of Consulting and Clinical Psychology*, 62, 75-82.
- Derogatis, L. R. (1993). *Symptom Checklist-90-Revised*. Minneapolis, MN: National Computer Systems.
- Downey, M. (1997). Trust me I'm a doctor *Sidney Morning Herald*, 10th May, 1.
- Eddy, D. K. (1996). Benefit language: criteria that will improve quality while reducing costs. *Journal of the American Medical Association*, 275, 650-657.
- Elkin, I. (1994). The NIMH Treatment of Depression Collaborative Research Program: Where we began and where we are. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of Psychotherapy and Behavior Change* (4 ed., pp. 114-139). New York: Wiley.
- Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). NIMH Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971-983.
- Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting*

- Psychology*, 16, 319-324.
- Flick, S. N. (1988). Managing attrition in clinical research. *Clinical Psychology Review*, 8, 499-515.
- Fonagy, P., Edgumbe, R., Target, M., Moran, G. S., & Miller, G. (2000). The Hampstead Manual of Child Analysis. London: Unpublished Manuscript: The Anna Freud Centre, London.
- Fonagy, P., & Higgitt, A. (1989). Evaluating the performance of departments of psychiatry. *Psychoanalytic Psychotherapy*, 4, 121-153.
- Fonagy, P., & Moran, G. (1993). Selecting single case research design for clinicians. In N. E. Miller, L. Luborsky, & J. P. Barber (Eds.), *Psychodynamic Treatment Research. A Handbook for Clinical Practice* (pp. 62-95). New York: Basic Books.
- Freud, S. (1933a). New introductory lectures on psychoanalysis. Lecture 35, The question of a Weltanschauung. In J. Strachey (Ed. & Trans.), In J. Strachey (Ed.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud (Original work published 1932)*. (Vol. 22, pp. 1-182). London: Hogarth Press.
- Group, E. B. C. R. (1994). Evidence based care 1. Setting priorities: how important is this problem. *Canadian Medical Association Journal*, 150, 1249-1254
- Grünbaum, A. (1984). *The Foundations of Psychoanalysis. A Philosophical Critique*. Berkeley Los Angeles London: University of California Press.
- Guyatt, G. H., Sackett, D. L., Cook, D. J., & Group, t. E. B. M. W. (1994). Users' guides to the medical literatures. *Journal of the American Medical Association*, 271, 59-63.
- Hellman, S., & Hellman, D. S. (1991). Of mice but not men. Problems of the randomised clinical trial. *New England Journal of Medicine*, 324, 1585-1589.
- Hope, T. (1995). Evidence based medicine and ethics. *Journal of Medical Ethics*, 21, 259-260.
- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Kagan, J. (1984). *The Nature of the Child*. New York: Basic Books.
- Kazdin, A. E. (1994). Psychotherapy for children and adolescents. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of Psychotherapy and Behaviour Change* (4 ed., pp. 543-594). New York: Wiley.
- Kernberg, O. F., Selzer, M. A., Koenigsberg, H. W., Carr, A. C., & Appelbaum, A. H. (1989). *Psychodynamic Psychotherapy of Borderline Patients* (B. dt (1993) Psychodynamische Therapie bei Borderline-Patienten. Hans Huber, Trans.). New York: Basic Books.
- Kerridge, I., Lowe, M., & Henry, D. (1998). Ethics and evidence based medicine. *British Medical Journal*, 316, 1151-1153.
- Klumpner, G. H., & Frank, A. (1991). On methods of reporting clinical material. *Journal of the American Psychoanalytic Association*, 39, 537-551.
- Kolvin, I., Garside, R. F., Nicol, A. R., MacMillan, A., Wolstenholme, F., & Leitch, I. M. (1981). *Help Starts Here: The Maladjusted Child in the Ordinary School*. London: Tavistock.
- Kukla, A. (1989). Nonempirical issues in psychology. *American Psychologist*, 44, 785-794.
- Lambert, M. J. (1976). Spontaneous remission in adult neurotic disorders: A revision and a summary. *Psychological Bulletin*, 83, 107-119.
- Lasch, C. (1978). *The Culture of Narcissism: American Life in an Age of Diminishing Expectations*. New York: Norton.
- Lilford, R., & Jackson, J. (1995). Equipoise and the ethics of randomizations. *Journal of Research in Social Medicine*, 88, 552-559.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York:

- Wiley.
- Luborsky, L. (1984). *Principles of Psychoanalytic Psychotherapy. A Manual for Supportive-Expressive Treatment*. New York: Basic Books.
- McNeilly, C., & Howard, K. (1991). The effects of psychotherapy: An reevaluation based on dosage. *Psychotherapy Research*, 1(1), 74-78.
- Morley, S. (1987). Single case methodology in behaviour therapy. In S. J. Lindsay & G. E. Powell (Eds.), *A Handbook of Clinical Adult Psychology* (pp. 699-718). London Gower Press.
- Morley, S. (1989). Single case research. In G. Parry & F. N. Watts (Eds.), *Behavioural and Mental Health Research: A Handbook of Skills and Methods* (pp. 233-264). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Parry, G. (1986). The case of the anxious executive. *British Journal of Medical Psychology*, 59, 221-233.
- Perron, R. (2001). Reflection on psychoanalytic research problems - A French-speaking view. In P. Fonagy, E. E. Jones, H. Kächele, R. Krause, J. F. Clarkin, R. Perron, A. J. Gerber, & L. Allison (Eds.), *An Open Door Review of the Outcome of Psychoanalysis. 2nd edition* (2 ed., pp. 3-9). London: International Psychoanalytic Association.
- Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin*, 108, 30-49.
- Rogler, L. H. (1989). The meaning of culturally sensitive research in mental health. *American Journal of Psychiatry*, 146, 296-303.
- Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine*, 53, 247-271.
- Roth, A., & Fonagy, P. (1996). *What Works for Whom? A Critical Review of Psychotherapy Research*. New York-London: Guilford Press.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71-72.
- Shapiro, D. A., Rees, A., Barkham, M., Hardy, G., Reynolds, S., & Startup, M. (1995). Effects of treatment duration and severity of depression on the maintenance of gains after cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, 63, 378-387.
- Shea, M. T., Elkin, I., Imber, S. D., Sotsky, S. M., Watkins, J. T., Collins, J. F., . . . , & Parloff, M. B. (1992). Course of depressive symptoms over follow-up: Findings from the National Institute of mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 49(10), 782-787.
doi:10.1001/archpsyc.1992.01820100026006
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The Benefits of Psychotherapy*. Baltimore: Johns Hopkins University Press.
- Spence, D. P. (1994). The special nature of psychoanalytic facts. *International Journal of Psycho-Analysis*, 75, 915-925.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-Trait Anxiety Inventory (Self-Evaluation Questionnaire)*. Palo Alto, CA: Consulting Psychologists Press.
- Stone, M. H. (1990). *The Fate of Borderline Patients: Successful Outcome and Psychiatric Practice*. New York: Guilford Press.
- Subotnik, L. (1975). Spontaneous remission of emotional disorder in a general medical practice. *Journal of Nervous and Mental Disease*, 161, 239-244.
- Webster, R. (1995). *Why Freud was Wrong: Sin, Science and Psychoanalysis*. London: HarperCollins.
- Williams, B. (1972). *Morality*. Cambridge: Cambridge University Press.

- Wilson, G. T., & Rachman, S. J. (1983). Meta-analysis and the evaluation of psychotherapy outcome: limitations and liabilities. *Journal of Consulting and Clinical Psychology*, *51*, 54-64.