The role of sampling in clinical trial design

MERTON KRAUSE, WOLFGANG LUTZ, & JAN R. BOEHNKE

Clinical Psychology and Psychotherapy, Department of Psychology, University of Trier, Trier, Germany
(Received 25 July 2010; revised 8 December 2010; accepted 13 December 2010)

Abstract
A treatment's recovery rate depends upon the percentage of clients who received the treatment and recovered. This rate is not logically interpretable as the personal probability of recovery of any individual client assigned to this treatment unless the rate is 0% or 100%. So clinical trials need to be designed to help us learn how to distinguish before treatment the sorts of clients who recover in response to each available form of treatment from those who do not. This requires our developing sufficiently comprehensive sampling of clients and client covariates as part of the design of clinical trials, which would be more likely and efficiently achieved were there centralized programmatic planning and coordination of the development of these aspects of clinical trial design.

Keywords: aptitude-treatment interaction research; outcome research; statistical methodology; client sampling; client-covariate sampling

The percentage of recoverers among the clients in the experimental treatment group of a clinical trial is often interpreted as if it were the personal probability of recovering for every single member of that comparison group. For example, if 60% of the clients in the experimental group of a clinical trial recover from a given malady, this result is not uncommonly reported as if each and every one of the clients in that group (and even of all similar persons with that malady) had a 60% probability of recovering if given the experimental treatment. If 40% of the clients in the no-treatment control group in this same trial also recover, this is not uncommonly reported as if each and every one of the clients in this trial (and even of all similar persons with that malady) had a 50% greater probability (60% being 50% more than 40%) of recovering if given the experimental treatment than if given no treatment.

Any scan of papers reporting controlled clinical trials or other experiments will show how common such inferences from sample percentages to personal probabilities (as well as to population relative frequencies) actually are. This continues to be so despite the past criticisms of this way of interpreting psychological data (e.g., Bakan, 1954; Estes, 1956; Merrill, 1931; Sidman, 1952; also Borsboom, Mellenbergh, & van Heerden, 2003; and, for some historical background, Danziger, 1990, pp. 68-153), including treatment: outcome data (e.g., Barlow, 1981; Cronbach, 1975; Kiesler, 1966; Strupp & Hadley, 1979; Yeaton & Sechrest, 1981). Better editorial policy and better training of researchers are needed if such misinterpretations are to be effectively discouraged.

In any given sample of clients with a given malady a percentage of recoverers that is between (and exclusive of) 0% and 100% for a given form of treatment leaves ambiguous what the expected results of this treatment are for subsequent individual clients with this malady. So long as clients are not distinguishable pre-treatment as either sufficiently like those who benefitted from the treatment or like those who did not, this ambiguity will remain. Such intermediate percentages of recoverers should serve primarily as an urgent invitation to researchers to try to find out what explains these differences, i.e., why a treatment succeeds with certain clients and fails with others. This calls for a new attitude toward within-comparison-group outcome variation: One that takes this variation not simply as a given fact but as a problem that needs solving, and then requires two demanding and centrally coordinated additions to the design of clinical trials for dealing with this within-comparison-group outcome variation.

Correspondence concerning this article should be addressed to Merton Krause, University of Trier, Psychology, Trier, Germany. Email: msk514@msn.com

ISSN 1050-3307 print/ISSN 1468-4381 online © 2011 Society for Psychotherapy Research
DOI: 10.1080/10503307.2010.549520
variation: a client-sampling plan for ensuring eventually comprehensive coverage of the population in need and a sample of client covariates for discovering how to adequately predict outcomes.

In order to focus as clearly and thus as narrowly as possible on this issue of within-comparison-group outcome variation due to outcome-relevant client characteristics, we shall consider here only the simplest possible model of a comparative clinical trial. It has one experimental treatment group, one no-treatment control group, and the two-valued outcome variable of recovery versus non-recovery. (More complex designs and finely graduated outcome variables certainly deserve consideration, but introducing them here would only distract from our purpose of making starkly clear why client and client-covariate sampling considerations are essential in clinical trial design.) We shall focus here on the design of clinical trials rather than on the analysis of their outcome data, because analysis cannot be better than design allows. In other words, data analysis methods for hedging against the danger of treatment contrasts being confounded by between-comparison-group differences in client composition are dependent for their effectiveness on client and client-covariate sampling being adequate. Discussions of such data analysis methods are readily available elsewhere (e.g., Kraemer, 2008, discusses covariance adjustment and e.g., Krause & Howard, 2002, Lutz et al., 2005, Ruberg, Chen, & Wang, 2010, Strobl, Malley, & Tutz, 2009, offer different approaches). So the principal issues to be dealt with here are why and how clinical trials ought to be designed in terms of the sampling of the client covariates and, in order to fully benefit from employing these covariates and to achieve adequate population representation, of the sampling of clients that these trials include in order for us to eventually learn how to adequately inform clinical decision-making about what sorts of treatments are best to provide to whom.¹

**Client Populations, Samples, and Individuals**

Each person i with a given malady can be considered to have an individual personal probability $P$ of recovering from it if treated in some given way $j$, $P$. For the two-valued outcome variable recovered versus not-recovered there is for each person with a given malady a personal probability of recovering if given treatment $j$ that is either 1 or 0, making $P$ equal either 1 or 0, because each person will in fact either recover or not recover. (Partial recovery is a complex notion in the theory of pathology that need not and cannot be done justice to here.)

It is obviously true that any given $n$-sized sample’s distribution of its individual members’ actual recovering or not recovering if given treatment $j$, i.e., the distribution of their $P$ values, determines the percentage of recoverers in that sample. This is because the mean $P$ of a sample’s $n$ members’ recovering equals the sample’s percentage of recoverers, i.e., the sum of all these personal probabilities of recovering (the ones and zeros) divided by $n$ equals the sample’s percentage of recoverers.

However, it is not true that a sample’s percentage of recoverers (so long as it is neither 0% nor 100%) determines or indicates anything about any individual member’s $P$ of recovering. For example, if the sample percentage of recoverers equals 50%, this can come about through any half of the members having $P$ values of recovering equal to 1 and the other half having $P$ values equal to zero. But just which of the sample’s members have $P = 1$ and which have $P = 0$ is obviously clinically important but is not at all determined or indicated when the sample has an intermediate percentage (>0 and <100) of recoverers.

So in order to distinguish pre-treatment between clients who will recover from some given malady because of receiving a treatment $j$ and those who will not, we must learn what client covariate (or there may be several) measurable pre-treatment is predictive of whether treatment $j$ will be effective. But because some clients may recover whether or not they receive a treatment $j$, we must also learn what client covariate (or there may be several) measurable pre-treatment is predictive of whether treatment $j$ is necessary.

**Untreated Recoverers and Treatment-Specific Recoverers**

**Untreated Recoverers**

One way a person with a given malady might be dealt with is by letting the malady simply take its natural course, i.e., by having nothing exogenous to the natural course of the malady intervene that would exacerbate or relieve the malady (Krause & Lutz, 2009a). In the whole population of persons with a given malady some will recover without treatment and some will not in the natural course of this malady for these persons. So for the first client covariate, and sticking to two-valued variables, let us call those who would recover untreated the “untreated recoverers,” $U$, and those who would not recover untreated the “not untreated recoverers,” $\sim U$.

Every clinical trial involves some sample of persons from the population of persons with the malady.
of concern for that trial. Perhaps some would recover untreated and so are untreated recoverers, \( \sim U \), while some do not and so are not untreated recoverers, \( \sim U \). Each client in any two-group randomized clinical trial’s (RCT’s) particular sample of clients is assigned by some random process either to a treatment group or to a no-treatment (ideally a natural course) control group. Therefore, in each of these two groups there probably will be some \( U \) and some \( \sim U \), although the percentages of these in the two groups most probably will be (a) unequal to their percentages in the population of persons with the given malady and (b) unequal to each other. Case (a) is so because clinical trials, whether RCTs or not, generally involve samples of clients who are merely somehow available to be included rather than representative or random samples of all the persons with the malady of interest. Case (b) is so because random assignment guarantees only a “statistically expected” equalization of these percentages as the average that is stochastically converged upon as the number of random assignments made from the given sample of clients approaches all the distinguishably different assignments possible from this given sample of clients (see Krause & Howard, 2003). However, in any given RCT there is only one such assignment made, and most of the mathematically possible assignments will necessarily yield unequal percentages of \( U \) and so, of course, also of \( \sim U \) in the two comparison groups. To see this, simply shuffle well a deck of playing cards, deal out two equal sized hands of cards that together exhaust the deck, and compare them. Repeat this operation as much as you please. The two hands will rarely be equal in color, suite, or face percentages, even though such equality is the “statistically expected” result. Each RCT involves just one such deal, just one such random assignment, and so this is obviously unlikely to produce two comparison groups equal in their percentages of \( U \).

**Treatment-Specific Recoverers**

Among the population of persons with a given malady, some will recover because they are given a particular treatment \( j \) and some will not (and likewise for any treatment \( k \) etc.). So for a second two-valued client-covariate, let us call those who would recover because of treatment \( j \) the “treatment \( j \) effectively treatable,” \( T_j \), and those who would not recover because of treatment \( j \) the “treatment \( j \) effectively untreatable,” \( \sim T_j \). In both the two comparison groups of an RCT that is meant to test the efficacy of treatment \( j \) there will probably be some \( T_j \) and some \( \sim T_j \). However, the percentages of \( T_j \) in the two groups will most likely be (a) unequal to the percentage of such persons in the population of persons with the given malady because of the sampling employed for obtaining clients and (b) unequal to each other despite and because of the random assignment of clients to the RCT’s comparison groups (as shown above).

Given these two covariates, \( U \sim U \) and \( T_j \sim T_j \), and given that every clinical trial starts with some particular sample of clients with a particular malady, the comparison groups in a two-group RCT involving this sample properly ought to adequately represent the population of persons with that malady in terms of the following four types of clients: Those who are effectively treatable by treatment \( j \) and would not recover if left untreated (\( T_j \& \sim U \)); those who are not effectively treatable by treatment \( j \) but would recover if left untreated (\( \sim T_j \& \sim U \)),\(^2\) those who are not effectively treatable by treatment \( j \) and would not recover if left untreated (\( \sim T_j \& \sim U \)), and those who are effectively treatable by treatment \( j \) but would recover if left untreated (\( T_j \& U \)).

For clinical purposes a sample is adequately representative if it includes some clients of every outcome relevant type that occurs in the population, but it need not include these in the same percentages as in the population.\(^3\) Such qualitative population representativeness of these types is essential if the reason for doing an RCT is to determine for what types of clients, i.e., for whom, with the given malady a particular treatment \( j \) is necessary and effective. This is so because all the recoverers in a treatment \( j \) group, who must be either \( T_j \& \sim U \), \( T_j \& U \), or \( \sim T_j \& U \), suggest prima facie that treatment \( j \) is effective for all these clients, because these are the sorts of clients who have recovered when given treatment \( j \). However, the \( U \) will recover without receiving treatment \( j \) and so what is clinically essential is that we learn how to distinguish the \( T_j \& \sim U \) from the rest if we are to determine for whom \( T_j \) is both effective and necessary. (See Pearl, 2000, pp. 173–200, for a fascinating and historically informative discussion relevant to this matter from a different perspective.)

**Learning how to Predict Untreated Recoveries and Treatment-Specific Recoveries**

The information that is needed for clinical purposes from studying the cases in RCTs is most obviously:

1. How to recognize and distinguish \( U \) and \( \sim U \) by comparing the clients who recovered with those who did not in the control group in order to construct a covariate upon which clients can be measured pre-treatment so that the \( U \) can avoid having to incur the costs of unnecessary treatment;

2. how to recognize and distinguish \( T_j \) and \( \sim T_j \) by
comparing the clients in the treatment \( j \) group who recovered with those who did not recover in order to develop a covariate upon which clients can be measured pre-treatment so that \( T_{js} \) can be spared the costs of what would be an ineffective treatment for them; (3) then, on the basis of (1) and (2) together, we can learn how to recognize \( T_j \) \& \( \sim Us \) pre-treatment, which is critically important because these and only these are the clients for whom some treatment is necessary and for whom treatment \( j \) is effective.

This will be most straightforward, however, only if the processes of treatment \( j \) and of natural-course recovery do not interact confoundingly for some clients and so when concurrently operative prevent the recovery of this type of client. If in the treatment \( j \) group of an RCT some \( Us \) fail to recover, these clients must be of this special type, and so they will be indistinguishable in treatment outcome from \( \sim T_j \) \& \( \sim Us \) but, unlike \( \sim T_j \) \& \( \sim Us \), this type of client does not need some treatment other than treatment \( j \) because they need no treatment at all. So in order to avoid having \( Us \) who are harmed by treatment \( j \) bias the development of a covariate for distinguishing \( T_{js} \) from \( \sim T_{js} \), it is necessary to identify and set aside any and all clients in the treatment \( j \) group of an RCT who qualify as \( Us \) (on the basis of the covariate developed or reaffirmed on this RCT's control group data for distinguishing \( Us \) from \( \sim Us \)) but fail to recover when given treatment \( j \). No such unrecovered \( \sim T_j \) \& \( Us \) or \( T_j \) \& \( Us \) should be included with the \( \sim T_j \) \& \( \sim Us \) when the covariate for distinguishing the recoverers from the non-recoverers in the treatment \( j \) group, i.e., for distinguishing \( T_{js} \) from \( \sim T_{js} \), is being constructed or tested. So only unrecovered \( \sim Us \) in the treatment \( j \) group, i.e., \( \sim T_j \) \& \( \sim Us \), rather than all the unrecovered clients should be compared with all the recovered clients of the treatment \( j \) group of an RCT in constructing or testing a covariate for distinguishing \( T_{js} \) from \( \sim T_{js} \).

This means that in doing post hoc analyses of RCT data the control group clients should be studied first for constructing or testing a covariate for distinguishing \( Us \) from \( \sim Us \) pre-treatment. This covariate should then be applied to the treatment \( j \) group clients in order to identify any \( Us \) who failed to recover when given treatment \( j \), and any such clients should be set aside for further study rather than be included with the \( \sim T_j \) \& \( \sim Us \) in constructing a covariate for distinguishing \( \sim T_{js} \) from \( T_{js} \) pre-treatment. Certainly if any treatment does conflict with rather than merely replace natural-course recovery for any clients, frustrating both sorts of recovery for them, it is clinically important to learn why this occurs. The \( Us \) who fail to recover when given treatment \( j \) (or \( k \) etc.), insofar as there are such, provide the basis for learning why.

It is the \( T_j \) \& \( \sim Us \) and only the \( T_j \) \& \( \sim Us \) that ethically should receive treatment \( j \), so long as there is nothing better to provide them (i.e., some treatment \( k \) that is somehow more effective or less costly, in non-monetary as well as monetary terms, for them than treatment \( j \)). This is the case because the key issue for clinical purposes is how best to treat each individual client, which means how to do what is both necessary and effective. How many and what percentages of clients there are for a given malady who are \( T_j \) \& \( \sim Us \), \( T_j \) \& \( Us \), \( \sim T_j \) \& \( Us \), and \( \sim T_j \) \& \( \sim Us \) (or \( T_k \) \& \( \sim Us \), etc.) is a public health rather than a clinical issue.

Thoroughly exploiting each RCT's potential for distinguishing the \( T_j \) \& \( \sim Us \) is something that is not as yet emphasized in RCT practice. To do so would require routinely maximizing the number of promising client covariates to facilitate the post hoc analyses necessary for seriously trying to distinguish the \( T_j \) \& \( \sim Us \) in each RCT's client sample, and then routinely doing numerous post hoc analyses and reporting their results. The constraint of significance testing has cast a pall on this essential post hoc work, which ought to discourage significance testing here rather than discourage this essential work (see Krause & Lutz, 2006; Ruberg et al., 2010), but this is too large an issue to get into here (for a start see, e.g., Krause, in press). Surely there can be nothing methodologically wrong with mining an RCT's outcome data with any number of candidate covariates so long as the findings are cast strictly as exploratory, especially when doing so is for serving the clinically crucial purpose of trying to determine just whom to treat specifically how.

**Traditional RCT Data Analysis**

The traditional design and analysis used in RCTs, however, do not serve this crucial clinical purpose but simply compares treatment group with control group relative frequencies of recovery (which for more finely than two-level graduated outcome variables obviously become two often overlapping distributions of relative frequencies over more than just two values). This traditional design and analysis makes the information for adjusting the treatment \( j \) group's percentage of recoverers used to compensate for the percentage of \( Us \) in the treatment group be the percentage of the \( Us \) in the control group. So the adjustment is made by subtracting the percentage of recoverers in the control group from the percentage of recoverers in the treatment group. But doing this can produce a correct estimate of treatment \( j \)’s relative frequency of actually needed effectiveness for all persons with the given
malady only if three assumptions hold: (1) Both the control and the treatment group accurately represent the percentage of Us in the population of persons with the given malady, and so both groups have the same percentage of Us; (2) the treatment group accurately represents the percentage of $T_i \& \sim Us$ in the population of persons with the given malady; (3) no Us in the treatment j group fail to recover.

If assumption (1) is not true (as is quite likely so long as we do not reliably recruit samples of clients for our RCTs that are representative of the relevant malady-stricken populations), then subtracting the percentage of recoverers in the control group, i.e., of Us, from the percentage of recoverers in the treatment j group will be incorrect as an adjustment of the latter percentage for the percentage of Us in the relevant population or in the treatment group. If assumption (2) is not true (as, for the same reason, is at present quite likely), then generalizing from the percentage of $T_i \& \sim Us$ in the treatment group to their percentage in the population of persons with the given malady is unjustified. If assumption (3) is not true, then the adjustment of the percentage of recoverers in the treatment j group on the basis of the percentage of recoverers in the control group fails to take account of any Us who fail to recover when given treatment j and so confuses some cases of non-necessity of any treatment with cases of ineffectiveness of treatment j.

If all these assumptions were to hold, there may appear to be some practical value for each clinician in knowing what percentage of all persons or of prior clients with a given malady are $T_i \& \sim Us$, $T_i \& \sim Us$, etc., because then he or she would have some basis for deciding which one treatment is the best bet for all his or her subsequent clients with that malady from among all the available treatments studied. How good a bet this is, however, depends upon the resemblance of the whole set of these subsequent clients to the relevant population of persons and to the relevant RCT's accumulated sample of prior clients with this given malady. Only the fact that the clinician cannot accurately predict pre-treatment which individual clients are $T_i \& \sim Us$, $T_i \& \sim Us$, etc. makes it necessary and worthwhile to take this gamble. And it is even a more risky gamble because no individual clinician's set of future cases of any given malady can reasonably be assumed to be well represented by their set of prior cases, by any single RCT's cases, by any accumulated set of RCT's cases, or by any population survey's sample of cases (were this last even possible) with regard to the percentages of $T_i \& \sim Us$, $T_i \& \sim Us$, etc. And nor can the present population of persons with any given malady be safely assumed to be well represented by traditional RCT's samples of such persons.

So the planful sampling of clients and of client covariates needs to be systematically included in the design of all clinical trials for distinguishing who is best treated by which treatment if clinicians are ever to reduce the riskiness of their bets on what course to take in treating each client. This is an RCT design issue that requires some central coordination if it is to be adequately and efficiently dealt with, e.g., coordination through funding agency policies, scientific journal requirements for accepting papers reporting RCTs or meta-analyses of them, and with government agencies and professional and scientific societies promoting, supporting, and actually organizing such coordination.

### Sampling and Clinical Trial Design

#### Client Sampling

How clients who have the malady that is being studied are sampled in an RCT that compares a treatment with a control condition is crucial for what the results of that RCT will be. In traditionally analyzed RCTs this is because clients of each of the four types, $T_i \& \sim U$, $T_i \& \sim U$, and $T_i \& U$ (as well as any subtypes of Us that fail to recover if given treatment $j$), are needed to be included in each comparison group of an RCT in proportions equivalent to their proportions in the population of persons with that malady. As we saw above, the greater the percentage of $T_i$ in the treatment j group (regardless of their percentage in the control group), the larger the percentage in the treatment j group of Us who are not blocked from recovering by being subjected to treatment j (regardless of their percentage in the control group), and the smaller the percentage of Us in the control group (regardless of their percentage in the treatment group), the more effective a treatment $j$ will appear to be in a traditionally analyzed RCT. These percentages depend jointly upon the particular sample of clients drawn from the population of persons with the given malady and the particular random assignment made from this sample to the RCT's comparison groups.

However, for clinical purposes the client sampling for a clinical trial of a treatment $j$ must simply be sufficient to ensure that each of these four types of clients is sufficiently included for covariate development or testing in the trial's comparison groups if the type exists at all in the population of persons with the malady of interest. These types' percentages in the treatment and control groups do not matter for clinical purposes so long as the types are sufficiently included for covariate development or testing when they are not zero in the population.
This is because for clinical purposes what is important is to learn from control group data simply how to recognize the \( U_s \) and then to learn from treatment \( j \) group data how to recognize the \( \sim T_j \) after any \( U_s \) who do not recover if subjected to treatment \( j \) are set aside. Only then can the \( T_j \& \sim U_s \) be identified in the treatment \( j \) group data in terms of the covariates by which they would be distinguishable pre-treatment from \( U_s \) and then from \( \sim T_j \). If these distinctions cannot be adequately made in a particular RCT, then the client sampling, the client random assignment, or the client covariate sampling were inadequate in that RCT and so it has failed to adequately serve any clinical purpose. (It can, however, still serve a research purpose by having included clients from some previously insufficiently studied subpopulation or provided useful information on some candidate covariates.)

To reduce the frequency of such failures, coordinated clinical trials are needed to improve the sampling of clients and client-covariates by pooling multiple diversely sampling RCTs' control group and, separately, treatment \( j \) group data in order to take advantage of their differences in client sampling and also to ensure that these RCTs' include as many as feasible of the same candidate client covariates. Population survey data on \( U_s \) may well provide information for recognizing \( U_s \) that is better than or importantly supplementary to what can be obtained from RCT control groups, as well as serving public health policy and epidemiological purposes.

**Client Covariate Sampling**

How candidate client covariates are sampled for trying to distinguish and recognize pre-treatment who is a \( U \), a \( T_j \& \sim U \), a \( T_k \& \sim U \), etc. is critical because we need accurate predictors of untreated recovery and of clients' differential treatability. We can only reasonably hope to discover such covariates through systematically trying out and subsequently testing numerous candidate client covariates. For clinical purposes if there is even a single person whose recovery when left untreated or whose differential treatability is not predictable by the current set of client covariates, then (1) additional such candidate covariates for distinguishing this type of client pre-treatment, (2) further client sampling to corroborate or further clarify the nature of this newly identified type of client, and perhaps (3) some new form of treatment that is effective for this (if \( \sim U \)) type of client is called for. All clinical trials, whether or not comparative or randomized, and so all effectiveness or evaluation studies and all ordinary practice, and also population surveys, can provide us important opportunities for taking promising client covariate measurements pre-treatment, for collecting historical data post-treatment or post untreated recovery or non-recovery on additional candidate client covariates, and doing the crucially important post hoc exploratory data analysis for helping us learn how to distinguish pre-treatment the \( U_s \), \( T_j \& \sim U_s \), \( T_k \& \sim U_s \), etc. This is something that at present is not being widely enough encouraged or done.

Because there are practical limits on how many pre-treatment measurements can be taken on the clients included in a clinical trial or in ordinary practice, only a relatively few candidate client covariates can be tried out in any single clinical trial's design or in anyone's ordinary practice. By centrally coordinating the addition to our stock of well studied candidate predictors of specifically treatment \( j \) or treatment \( k \) etc. recovery (i.e., of who is a \( T_p \), \( T_k \), etc.) and of untreated recovery (i.e., of who is a \( U \)), it should be possible to progressively eventually develop a predictor battery (i.e., one with an \( R^2 \) that eventually approaches 1) for identifying \( U_s \), \( T_j \& \sim U_s \), \( T_k \& \sim U_s \), etc. Preferably all client covariates that have already shown some relevant predictive promise need to be reliably incorporated along with further plausible candidates in all subsequent clinical trials as well as in ordinary pre-treatment assessment procedures (where, ideally, the same outcome variable or battery would also be universally used). This needs to go on until covariate batteries sufficiently differentially predictive of recovery, i.e., of recovery when treated this or that way or untreated, have been well evidenced. For this to be accomplished most economically would require some central coordination across clinical trials of the client covariate sampling features of trial design and across clinical settings of pre-treatment assessment. It would also require central coordination of client sampling to ultimately ensure for clinical purposes qualitative population representativeness across clinical trials and across clinical settings for each malady. For public health purposes this would also require achieving accurate estimates of population frequencies of \( U_s \), \( T_j \& \sim U_s \), \( T_k \& \sim U_s \), etc.

Suggesting a list of plausible or empirically promising candidate client covariates properly requires extensive theoretical work and a major investment in literature reviewing that exceed the ambitions of the present paper, but certainly some personal attributes are possibly promotive of untreated or of variously treated recovery from mental illness. These include personality traits such as ego strength and intelligence; informal social relationships that are, e.g., honest, empathic, and positively regardful; socio-economic advantages such as affluence and status;
attitudes toward one's personal involvement in psychotherapy such as motivation for and understanding of it; and mental health status itself (as to this last see Cook & Steiner, 2010; Driessen, Cuijpers, Hollon, & Dekker, 2010; Luborsky et al., 1993). Of course we must first of all have consensually valid measures for such client covariates.

Comparing Different Treatments' Effectiveness

In general not all clients with a given malady respond favorably to any given form of treatment for that malady. So it is important for clinical purposes to be able to accurately predict which available form of treatment will lead to recovery for each different outcome-relevant type of client and so for each individual client. For any given client this may not be the treatment that is the most relatively frequently effective one, because how relatively frequently effective a treatment is for any given group of clients depends upon the composition of that group. For some clients with a given malady a treatment \( j \) may be effective but a treatment \( k \) ineffective \((T_j & \sim T_k)\), for some others it may be the other way around \((\sim T_j & \sim T_k)\), for still others both may be effective \((T_j & T_k)\), and for the rest neither may be effective \((\sim T_j & \sim T_k)\). For example, in an RCT comparing treatments' \( j \) and \( k \), the sample of clients with the malady of interest might be composed of 200 \( T_j \), 100 \( T_k \), 50 \( \sim T_j \) & \( T_k \), and 100 \( \sim T_j \) & \( \sim T_k \). If the random assignment of these clients to the two comparison groups happened to result in precisely these same proportions of 4: 2: 1: 2 in both these groups, then treatment \( j \) will appear 1.2 times as effective as treatment \( k \) in this RCT. For clinical purposes and whatever the sample size, it would obviously be a mistake to certify treatment \( j \) and decertify treatment \( k \) as appropriate to use in all cases of this malady merely on the basis of which of them was relatively more effective in this particular RCT. And a meta-analysis of a set of RCTs comparing the same two treatments is simply uninterpretable so long as the \( T_j \) & \( T_k \), \( T_j \) & \( \sim T_k \), \( \sim T_j \) & \( T_k \), and \( \sim T_j \) & \( \sim T_k \) proportions in the included RCTs' comparison groups vary in unknown ways over these RCTs, as is currently likely to be the case.

A treatment's failure to outperform another treatment in the aggregate does not logically mean that the former treatment ought simply to be wholly abandoned, because a treatment's not bettering another treatment in the aggregate does not mean that it does not better the latter for some sorts of clients (see Howard, Krause, & Vessey, 1994). So \( T_j \) & \( \sim T_k \) must be distinguished on the basis of covariates from \( \sim T_j \) & \( T_k \) in order to optimize treatment assignment. \( T_j \) & \( T_k \) must be distinguished from the rest in case treatment \( j \) is simply less costly than treatment \( k \) or vice versa. \( \sim T_j \) & \( \sim T_k \) must be distinguished from the rest as the basis for trying to discover still other treatments specifically for such clients. In fact treatments should be certified as effective enough for and only for the specific sorts of clients for whom they are at present evidently reliably cost-effective enough. All clients ethically ought to be informed about any present uncertainty of the cost-effectiveness of such treatments for them so they can give or withhold their informed consent, if the requirement of informed consent is to be taken seriously.

Conclusions

Until we can identify pre-treatment who with a given malady would and who would not adequately benefit from each available form of treatment, the percentages of recoverers in the different comparison groups of an RCT involving such clients merely provide unreliable betting odds for choosing treatments for individual clients in practice. In other words, if one form of treatment has so far resulted in a higher percentage of recoverers than another and there is no way to predict which recipient of either form will recover and which will not, then the only guidance a clinician can draw from this difference in recovery rate is that the treatment form with the higher percentage of recoverers may be the better bet for all clients with that given malady. The more this percentage deviates from 100% and 0% the poorer its guidance, because so long as some clients are not adequately benefited by the treatment with the highest percentage of recoverers or some would benefit uniquely from the treatment with the lowest percentage such a bet will sometimes fail for them. And the percentages of clients in a clinician's case load who would benefit from each alternative treatment may differ considerably from the percentages in the relevant RCTs.

Only when we have eventually included in RCTs all the outcome-relevant types of clients with a given malady and all the client covariates necessary for distinguishing pre-treatment among all these types, and the treatment comparisons are made separately for each type of client, can RCTs be most useful for clinical purposes. For public health purposes the frequencies in which these types occur in the general population also need to be known. Clinicians' pre-treatment assessment of their clients in terms of the prognostic client covariates that distinguish these types would then indicate what form of treatment is evidently best for each client. So it is essential to RCT design that we add client sampling plans for ensuring comprehensive coverage of all the
outcome-relevant client types in the population of persons with the malady at issue, and that we add candidate client covariate sampling plans for systematically developing a uniform set of client covariates for distinguishing these types. This needs to be done in a centrally coordinated manner.

Although psychotherapy has been the focus of this paper, the logic of the argument advanced here applies to clinical medicine. With only a little adjustment of terminology it also applies to teaching, parenting, managing, and all the other psychological crafts and professions.

Notes

1 The essential clinical purpose in doing two-phase (input-outcome) RCTs is to resolve the issue of to what treatment it is best to assign each client. Most RCTs have so far been two-phase rather than multi-phase (input-process-outcome) studies. To resolve the equally important issue of how the therapist ought best to manage the next phase of treatment in the light of how the client has so far responded is the essential clinical purpose in doing multi-phase input-process-outcome studies, which cannot be done justice to here (see, e.g., Davison, 2000; Krause & Lutz, 2009b; Lambert et al., 2002; Lutz et al., 2009; Ruberg et al., 2010).

2 There is the logical possibility of there being persons who would recover if left untreated but not if given treatment j, and so an alternative kind of U, but this may not so obviously be an empirical possibility. For those who prefer the latter and so at least some plausible example, the following hypothetical situation (and also Barlow, 2010) may be useful. Perhaps for some persons after a severe psychological trauma, enough time spent in non-stressful circumstances, and without any felt prospect of being subsequently subjected to stressful circumstances, there occurs an adequate diminution of post-traumatic stress disorder (whether this is by gradual forgetting, consistent extinction trials, repression, or however). Such a person would be a U. However, if for such persons the prompted remembering of the circumstances of the trauma (or of events in her or his life preceding it that might have contributed to making these circumstances so traumatic) interferes with this natural course of diminution, then this stressful intervention treatment (a treatment j) may prevent the prior sort of recovery but may not itself produce recovery. Such a person would be a U who when given treatment j would fail to recover because the treatment interfered with the processes of natural recovery and either provided no effective substitute for them or was itself impeded by them. If this treatment did provide such an effective substitute for and also was not impeded by such natural processes of recovery, then the person would be a Tj&U.

3 Whereas knowing these percentages and their underlying frequencies is crucial for public health purposes. Public health policy concerns the investment of public resources so as to minimize the social cost of ill health, which necessarily requires consideration of how many clients of what sorts have what maladies, how these numbers are trending over time and place, what treatments (including simply not intervening) are cost-effective enough for what sorts of clients with which of these maladies, and for what maladies and sorts of clients are new treatments most surely needed to control the social cost of illness. If (a) enough persons do or are likely to suffer from a particular malady and their doing so has a social cost exceeded by few enough competing social problems in the light of society's present tolerance for social problems and resources for dealing with them, (b) a large enough percentage of these persons recover from this malady after receiving a particular treatment j who currently would not otherwise (Tj& ~ Uj), Tj& ~ Tj&j, etc.), and (c) the net marginal impact on total social costs of providing treatment j is favorable or at least socially tolerable, then investing public health resources in treatment j is socially worthwhile. Population surveys of the prevalence, incidence, outcome, and social cost of each malady when left untreated and also clinical outcome surveys of ordinary practice (rather than merely RCT estimated) cost-effectiveness of alternative available treatments for each malady are necessary for rationally developing public health policy, because only in the light of such information can clinical trials of new treatments for particular maladies be rationally prioritized for public health resources' investment. Thus, considerable ambiguity as to the expectable cost-effectiveness for individual clients is (and simply has to be) tolerable for public health purposes but properly is not tolerable for clinical purposes. With client and client covariate sampling taken seriously the traditional RCT would be a proper tool for informing both public health and clinical policy.

4 This ought not to be presumed to be simply a correlational matter, with each full-rank covariate correlating with one or more outcome variables. It is also possible that configurations of several covariates' levels are associated with particular levels (or configurations of levels) of the outcome variable(s) even though none of the individual predictor variables correlates at all with the outcome variable(s). In other words, the covariates may be quite irregularly associated with or interactive in their effects on outcomes (see Krause, 2010; Krause & Howard, 2002) and so are a matter of non-linear association rather than linear correlation (see Lutz et al., 2005; Ruberg et al., 2010; Strobl et al., 2009; West & Thoemmes, 2010).

5 Some comment is deserved on the issue of counterfactual or potential responses (see, e.g., Flanders, 2006; VanderWeele & Hernán, 2006; West & Thoemmes, 2010), in that each client receives one particular treatment and not another, and so there is no way to tell empirically whether any given client would have recovered if that client had received the other treatment. This logical obstacle is routinely finessed statistically by making the assumption of sufficient relevant similarity (uniformity, exchangeability) between the groups of clients who receive different treatments (see, e.g., Dawid, 2000), which assumption holds at the mathematical limit (i.e., as a "statistically expected value") on the basis of randomly assigning to its comparison groups whatever clients are available for inclusion in an RCT but is unjustified for any single RCT in the analysis. Whatever clients are available for inclusion in an RCT but is unjustified for any single RCT (Krause & Howard, 2003; West & Thoemmes, 2010, p. 23). So, instead of focusing on individual clients, the logic of the RCT has focused on groups of clients, on each group's mean outcome. However, in order logically to bring the benefits of this maneuver down to the individual client level, it is necessary to consider distinct types of clients so that each client of each particular type responds to any given treatment in the same way that every other client of that type does, e.g., either all those of a particular type recover or none does. Covariates are needed to predictively distinguish the various such outcome-homogeneous types of clients if clinical decision-making is to be thoroughly benefited from such uniformities of response by client type. Such perfectly outcome-homogeneous types of clients are thus the logical stand-in for individual clients and would allow us to make unambiguous inferences about individual clients from clinical trial data, which does not require potential response theory to see but is seen by those who are working at such theory (see West & Thoemmes, 2010).
Sampling and clinical trial design

References


Cook, T.D., & Steiner, P.M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcomes, of unreliable measurement, and of mode of data analysis. Psychological Methods, 15, 56-68.


