1

Research Methods

Julian A. Rubel and Wolfgang Lutz

Introduction

In most areas of psychology, chapters on research methods are predominantly concerned with the description of well-controlled conditions of laboratory studies and their proper analysis. However, the scope of clinical psychology is much broader than that of basic psychological science and laboratory studies. The variety of topics ranges from foundational issues to applied contexts. As clinical psychology is a far-reaching field of applied psychology, much research is concerned with phenomena that could not easily be studied in the lab or under controlled conditions. As a consequence, research methods within clinical psychology need to include designs and evaluation strategies ranging from laboratory studies to clinical interventions as they are delivered in the field. However, instead of making considerations about research methodology less important, this broader focus increases the importance of a knowledge of methodological issues to allow the appropriate analysis and interpretation of study results (Kazdin, 2013). Increased sophistication of applied research methods helped clinical psychology to establish itself as a profession. Regardless of their future occupation, a firm understanding of research methods is pivotal to every scholar in clinical psychology. Clinical scientists must not only be acquainted with research design considerations and statistical concepts, they also need to have expertise in this area to be able to provide a treatment that is based on scientific evidence.

The present chapter provides a nontechnical overview of the most important concepts of research methods in clinical psychology. In the first section of this chapter, central concepts pertaining to the study of the frequency, development and prevention of psychological problems are described briefly. Since most research in clinical psychology is on interventions, the second part of this chapter deals with the evaluation of these treatments. In this section, we present methods that are concerned with the following three overarching questions: (a) Does the intervention work? (b) Is the intervention effective for a specific patient? (c) How, for whom, and under which conditions does the intervention work?

Research on the Frequency, Cause, and Prevention of Psychological Problems, and Disorders

Epidemiology

Much research within clinical psychology attempts to answer questions such as: Who has a psychological problem or disorder? How is a disorder distributed in a specific population? Which factors lead to or increase the risk of psychological disorders? How does an untreated disorder develop? Who is seeking treatment and who needs it? The field of *epidemiology* deals

with these questions (e.g., Rockett, 1999). *Descriptive epidemiology* deals with the distribution (occurrence, spatial, temporal) of these phenomena, and *analytic epidemiology* deals with the determinants (causes) of psychological disorders. Important concepts in epidemiological research are described below.

Prevalence

Prevalence indicates the frequency of a psychological disorder, generally or in a specific population. The prevalence rate is the proportion of people with a specific disorder in relation to the population of interest. Prevalence must be specified with regard to a particular time period and the examined population: For example, *12-month prevalence* refers to the rate of occurrence within a period of 12 months. In comparison, *lifetime prevalence* refers to the entire lifespan. Instead of a time period, prevalence can also refer to a specific time point (*point prevalence*). An additional important figure is *treatment prevalence*, which is not concerned with the frequency of occurrence of a disorder but the frequency with which persons seek treatment for a specific disorder.

Incidence

Incidence refers to the number of persons in a given time period and population that newly develop a disorder. Thus, the incidence rate is the proportion of persons in a given population that have a disorder but did not have that disorder in the past. In accordance with this definition, two measurement points would be necessary for a valid incidence estimate: The first time point provides the base-rate of people in a population who do not suffer from the disorder. The second time point determines the number of patients who were not ill at the first time point but are ill now. Like prevalence, incidence depends on the investigated period, and the population. If, for example, the second measurement point is one year after the first measurement point, the incidence rate is specific for this 1-year period.

Risk Measures

Generally, two types of risk measures can be differentiated: *unconditional risks* and *conditional risks*. *Unconditional risks* address the likelihood of developing a specific disorder in a given period. These risks can be calculated with the respective prevalence and incidence estimates described above. *Conditional risks* address whether certain variables increase (risk factor) or decrease (protective factor) the probability of developing a disorder. As such, whether the prevalence and/or incidence rates differ is investigated depends on the variable in question (e.g., sex). Many psychological disorders occur more frequently in women than in men. Consequently, being female is a risk factor for the development of these disorders.

Etiology and Analytical Epidemiology

When investigating the causes of psychological disorders, multidimensional models are usually assumed. That is to say, psychopathology is too complex to be explained by a single cause. Rather, many different influence factors from multiple dimensions are thought to interact, and eventually result in a psychological disorder. Etiology and analytical epidemiology address the questions of *who* develops a disorder and *under which circumstances*, taking into account behavioral, biological, emotional, social, and developmental influences. To observe the relative influence of each of the different factors, similar methods are applied, as described below (also see the section on the *control-group experiment*). The basic idea is to investigate groups that differ with regard to certain influence factors and are identical with regard to others. The examination of the effects of genes, for example, is often done within so called "twin studies."

Twins are identical with regard to their genetic code but might be exposed to other very different influence factors, especially if they were raised apart from each other. Those characteristics, which are shared by twins after many years within different environmental conditions, are highly likely to have strong genetic influences.

For the design of examinations that seek to establish causal influence factors, it is important to show that the potential influence factor was present before the disorder. Therefore, the repeated assessment of the same individuals over time is needed (*longitudinal designs*). *Cross-sectional designs*, in which data is collected from different age groups at the same time, can also hint at causal associations. However, this design assumes that the age groups are comparable with regard to other, not measured characteristics. If there are systematic differences between the different age groups (*cohort effects*) these can hamper the interpretation of cross-sectional studies.

Prevention

Besides the treatment of psychological disorders, the prevention of their onset is crucial for clinical psychology. Prevention research within clinical psychology investigates interventions or programs that help to reduce the risk of developing a psychological disorder. While *primary prevention* programs aim at risk reduction on a global level (e.g., for all inhabitants of a country), *secondary prevention* focuses on individuals who already show an increased risk of developing a disorder or already report subclinical problems. As such, prevention research is based on etiology and epidemiology, as knowledge on the potential causes of psychological disorders is needed to create effective programs. The evaluation of these programs uses the same methods as those presented below for the evaluation of other clinical interventions.

Evaluating Clinical Interventions and Treatments

Central to clinical psychology is the question of the effectiveness of specific clinical interventions as well as complete psychological treatments (e.g., cognitive behavioral treatments, psychodynamic treatments). The first step in the process of evaluating psychological interventions and treatments is an appropriate definition of the program or intervention, and the identification of criteria that differentiate success from failure. In psychotherapy research, for example, it is agreed that assessments of outcomes should not be limited to a single dimension (e.g., depressive symptoms), even if the focus of the study is a specific disorder (e.g., depression). While symptoms should be one of the primary outcomes, most studies collect data along multiple dimensions (e.g., work/social adjustment, interpersonal problems etc.), and include different perspectives (e.g., patient ratings, therapist ratings, third-party ratings). While psychophysiological and neurocognitive procedures have recently emerged as a new way of measuring change, questionnaires are still the predominant method of choice (e.g., Ogles, 2013).

In clinical studies, these outcome criteria are used as *dependent variables* (DV), which are assumed (*hypothesized*) to differ between persons depending on one or more manipulated or observed *independent variables* (IV). The most common IV in clinical research are interventions. If a researcher hypothesizes that 6 weeks of an intervention A are more effective in reducing symptoms of depression than just waiting 6 weeks, patients would be assigned into two groups: One group would receive intervention A, the other would not. Thus, these groups differ with regard to the IV *treatment* (*intervention A versus waiting*). If, after the 6 weeks, patients who underwent the intervention show less depressive symptoms (DV) than those in the waiting group, the researcher's hypothesis is confirmed. However, it must be ensured that

there are no alternative explanations for the differences in the DV other than the difference in the IV (intervention versus waiting). For potential threats to this causal interpretation and means of ruling out competing explanations, see the sections on *internal* and *external validity*.

In clinical psychology, the aim is often the amelioration of relevant symptoms. A crucial task in clinical research is therefore the measurement of change. "Measuring" denotes the determination of patients' characteristics regarding specific attributes. With regard to the measurement of change, two types can be differentiated: *Retrospective* and *repeated assessments change measurement*. *Retrospective change measurement* uses retrospective ratings of the amount of change induced by an intervention. This can be realized via global success ratings at the end of the treatment or by questionnaires specifically developed for this purpose. *Repeated assessments change measurement* uses differences in the scores from ratings at the beginning and the end of the intervention. Both approaches have specific advantages and disadvantages. *Retrospective* measurements allow an immediate and economic assessment of change. However, this approach enables no objective comparison with the state at the beginning of the treatment. *Retrospective* estimates are prone to several biases, which are typical of retrospective ratings. *Repeated assessments* rely on the principles of *classical test theory* (CTT). Since CTT struggles with an appropriate conceptualization of change measurement, related issues apply to *repeated assessments* (e.g., the problem of regression to the mean, the reliability of difference scores and the stability of the construct over time; for an in-depth discussion of these issues refer to Crocker & Algina, 1986). In clinical studies, multiple assessments have become standard.

Having defined appropriate criteria for the description of a course of change, the question arises of *when* these criteria should be assessed and after what amount of time an intervention can be considered successful. In order to test the *stability* of effects, the observed change must remain stable after termination of the intervention. Thus, conducting the last assessment at the end of an intervention cannot be enough to confirm its effectiveness. Instead, in order to be able to assess the stability of effects, the evaluation design must include measurements that are timed several weeks, months or even years after the end of treatment (follow up).

Does the Intervention Work?

To establish the effectivity of an intervention, it is crucial that the observed change can be attributed to the intervention with certainty. That is to say, alternative explanations of this change must be eliminated. To rule out as many alternative explanations for the observed change as possible, the *control-group experiment* has been considered the "gold standard" in clinical research. In *control-group experiments*, patients are randomly assigned to the intervention or a control condition. The objective of random assignment is the complete interchangeability of the groups before the start of the experiment. If this is achieved, every difference between the groups that is observed after the experiment can be attributed to the difference between the intervention and the control condition (manipulated IV). To be able to draw very specific conclusions, the difference between the intervention and the control condition should be limited to the specific factor that is hypothesized to cause the effect of the intervention. All other factors should be kept constant. These kind of studies are called "randomized control(led) trials" (RCTs). Depending on the respective control condition, different conclusions can be drawn (see Table 1.1).

Internal Validity

Randomized controlled trials aim to test hypotheses deductively, for example with respect to the effectivity of a newly developed intervention in comparison to an established intervention.

Table 1.1 Different control groups and potential corresponding study conclusions.

Control condition	Potential conclusion
Waitlist control (participants receive no intervention and are just assessed before and after the experiment; after the experiment these participants receive treatment)	Intervention A is more effective than no intervention
Alternative intervention A—without effective ingredients (participants receive a placebo treatment)	The effects of intervention A are not only due to a placebo response
Alternative intervention B—with other effective ingredients (participants receive a different intervention, which is assumed or has been shown to be effective)	Intervention A is more effective than intervention B

Thus, aspects of *internal validity* are emphasized. *Internal validity* describes the certainty with which the observed differences between the experimental conditions can be attributed to the manipulations in the experiment (i.e. the clinical-psychological intervention). As described above, ruling out alternative explanations is key to this approach. In clinical psychology, the following measures are often taken to secure internal validity:

- *random assignment to the conditions* to secure the comparability of the groups and rule out person characteristics as alternative explanations;
- *homogeneous samples (i.e. clearly specified diagnostic groups)* to draw specific conclusions for specific populations;
- *a strict standardization of the intervention (e.g., by manualization of the intervention)* to ensure that the intervention is conducted as intended for every participant—this regularly includes post hoc assessments of protocol adherence and the competence with which the protocol was implemented;
- *training of those who conduct the intervention* to ensure a comparable competence of the therapists.

External Validity

Despite the methodological rigor of experimental clinical research, it is repeatedly criticized for its narrow emphasize on internal validity, which is often achieved at the cost of external validity (e.g., Howard, Moras, Brill, Martinovich, & Lutz, 1996). *External validity* describes the possibility of transferring study results to practice settings and is emphasized in quasi experimental or naturalistic studies. The transport of evidence from the lab to the field of clinical psychology (i.e. everyday clinical practice) represents a separate and important issue, and involves questions of generalizability, feasibility, and cost-effectiveness of therapeutic interventions. Quasi experimental studies aim to investigate the extent to which interventions are effective in clinical practice, without the controlled conditions of an RCT. This is important, because strict selection criteria in experimental studies with regard to both participants (i.e. homogenous sample) and therapists (i.e. specifically trained therapists) may limit the generalizability of results. Instead of the a priori control of potential confounding variables, the results of naturalistic studies are often controlled post hoc using statistical methods (e.g., ANCOVA). Unfortunately, the relationship between internal and external validity is reciprocal. Consequently,

Table 1.2 Basic differences between naturalistic / quasi experimental and experimental clinical studies.

Naturalistic / quasi experimental studies	Experimental studies
Explorative/inductive	Confirmatory/deductive
External validity	Internal validity
Heterogeneous samples	Homogeneous samples
Nonmanualized treatment	Protocol-based treatment
Statistical control	Randomization

the focus on external validity in naturalistic studies comes along with potential threats to internal validity, which could hamper clear-cut or even causal conclusions. Table 1.2 depicts the major differences between experimental and quasi experimental studies (e.g., Lutz 2002).

Given the divergent foci of these two types of research, it is commonly accepted that evidence from both are needed to build a solid evidence base for an intervention. This is also emphasized by the separate terminology introduced for both kinds of investigations: While naturalistic studies can validate the effectiveness of a treatment (*effectiveness studies*), randomized controlled trials test the efficacy of a treatment (*efficacy studies*).

Quantifying the Effects of an Intervention

Effect Sizes

With the help of effect sizes, it is possible to compare the results of different studies and estimate differences in effects between different conditions (e.g., psychotherapy versus waiting-list control group or comparisons between different interventions). Effect-size measures allow a quantification of differences between studies, irrespective of the applied instrument. There are different ways to calculate effect sizes. Generally, these can be divided into effect sizes that compare two different groups (e.g., treatment vs. control) and effect sizes of within-group comparisons (e.g., pre-post comparison). Basically, these types of effect sizes are differences of the scores of the compared groups (IG scores minus CG scores after the treatment or preintervention scores minus postintervention scores) in a specific instrument (e.g., a measure for depression). To be able to compare these effect sizes between different instruments, these differences are standardized at (i.e. divided by) the amount of variation in the respective scores (i.e. standard deviation; SD).

The literature discusses which standard deviation should be used to standardize the difference scores—the SD of the prescores/CG, the SD of the postscores/IG, or a pooled SD taking into account the variation at both time points or in both groups. Although the different techniques to calculate effect sizes produce similar results, they might lead to substantial differences.

The effect sizes described so far belong to the so-called d-family (e.g., Cohen's d or Hedges' d) and are calculated based on standardized average differences between two populations. Besides the d-family, the r-family is a prominent effect size measure in the literature. The calculation of effect sizes in the r-family is based on correlations (r). A correlation is a measure of the common variation of two variables and ranges from −1 to +1. Positive correlations ($r > 0$) between two variables A and B indicate that increases in A go along with increases in B and decreases in A go along with decreases in B. High negative correlations ($r < 0$) indicate that increases in A go along with decreases in B and decreases in A go along with increases in B. A correlation of 0 indicates that A and B are completely unrelated. The squared correlation (r^2) is called the

determination coefficient and denotes the share of the variation in B or A that can be explained by A or B respectively. That is to say, r^2 tells us something about the percentage of the differences in a variable A that can be explained by a variable B. For example, in treatment research it would be interesting to know how much of the differences in the patients' postscores can be explained by the treatment variable (i.e. treatment A or treatment B). Depending on the characteristics of the investigated variables (both continuous, both categorical or mixed) different kinds of correlations can be calculated, however their interpretation generally remains the same.

Clinical Significant Change

While the effect-size measures described above allow the quantification and standardization of group differences, they do not allow conclusions with regard to the clinical significance of the observed change. Several concepts have been developed to determine the clinical significance of measured changes. The most commonly applied concept of clinical significant change is described briefly below (e.g., Jacobson & Truax, 1991). Jacobson and Truax's approach provides a statistical criterion that allows the determination of the amount of change that could be considered as clinically relevant for each patient. This concept is composed of two conditions: (a) Change from pretreatment to posttreatment must be reliable (i.e. likely not a mere consequence of random variation and measurement error) and (b) a patient's score after treatment must have a higher probability of belonging to healthy sample than to a distressed sample. To calculate the amount of change that can be considered reliable (or statistically significant), the pre-post difference is related to the measurement error of the applied instrument. The minimal amount of change considered to be reliable is the *reliable change index* (RCI):

$$RCI = 1.96 \times \sqrt{2 \times (SD \times \sqrt{1-r})^2}$$

where *SD* reflects the standard deviation of a reference sample and *r* the reliability (e.g., internal consistency) of the respective instrument in a similar sample.

For the determination of a cutoff score, Jacobson and Truax (1991) suggest three different options, depending on the available reference data for the applied instrument. Criterion A: Only data from a clinical reference sample is available. The cutoff score (C_a) is defined as two standard deviations ($2*SD_{clin.}$) below the mean of a clinical reference sample ($M_{clin.}$). Criterion B: Only data from a healthy reference sample is available. The cutoff score (C_b) is defined as two standard deviations ($2*SD_{nonclin.}$) above the mean of a nonclinical reference sample ($M_{nonclin.}$). Criterion C: Data from a clinical and nonclinical reference sample are available: The cutoff score (C_C) is defined as the value, which is equally likely to stem from the clinical or from the nonclinical sample.

Integrating the Results from Multiple Studies—Meta-analyses

Replication and Stage Models

Which of the designs described above is more appropriate for the evaluation of clinical interventions has been debated (e.g., Howard et al., 1996). There is consensus that a single study is not sufficient to consider an intervention as evidence based. Rather replications (repeated investigation of the same research question) in experimental as well as naturalistic settings are needed. Similar to medical research, a stepwise approach has been proposed for the evaluation of clinical interventions. The National Institute of Mental Health developed a stage model for testing a new treatment program, which stipulates clinical-experimental studies in stages 1 and 2. In stage 3, the generalizability and feasibility should be tested in quasi experimental

studies (e.g., Rounsaville et al., 2001). It is important to note that different designs answer different questions that all are important for a comprehensive evaluation of an intervention: experimental studies allow conclusions concerning the effect of a specific (e.g., newly developed) treatment approach under controlled experimental conditions. Quasi experimental and naturalistic studies give information about the generalizability of the results to clinical practice.

Meta-Analyses

The results of individual studies provide only limited information concerning the effectivity of an intervention. Given the problems of individual studies described above, an observed effect can always be a result of the selection of specific subjects, specific settings or designs, or other reasons (e.g., differences between the therapists in the experimental conditions). This is true both for results from naturalistic studies and RCTs. Therefore, replications are a pivotal threshold for all clinical-psychological research results. One possibility to integrate the results of several studies investigating a specific intervention is meta-analysis (e.g., Hunter & Schmidt, 2004). The aim of a meta-analysis is to summarize studies focusing on a particular research question and aggregate them statistically. The effect sizes reported in the original studies are extracted. As described above, an effect size can be, for example, a correlation (r) or a standardized mean difference (d). In a next step, these effect sizes are aggregated for all studies that are included in the meta-analysis, i.e. an overall effect size is calculated. Given the fact that meta-analyses are based on the results of several original studies, conclusions can be derived on a broader evidence base. However, in order to interpret the results of meta-analyses it is important to note that they might rely on very different outcome measures and heterogeneity with regard to the realizations of the treatments (i.e., "comparing apples and oranges"). Another important criticism concerns the dependency of meta-analyses on the quality of the original studies ("garbage in, garbage out"). Therefore, new meta-analyses often take into account the quality of the included original studies (e.g., design, sample size, etc.) as well as differences in the results for different outcome measures (e.g., Hunter & Schmidt, 2004). These, as well as other potentially influencing variables, can be obtained for each study and tested with regard to their effects on the overall effect size estimate in so called *metaregression analyses*. If data on individual patient characteristics that might influence the overall effect size estimate are available, *individual patient-level meta-analyses* might be a more reliable option to analyze moderating effects.

Recently an extension of conventional meta-analyses found its way into clinical research, namely *network meta-analyses*. While conventional meta-analyses only allow inferences about treatments that were directly compared to each other in the original studies, network meta-analyses also allow indirect inferences (e.g., Lumley, 2002). That is, if many original studies directly compared treatment A with treatment B as well as treatment B with treatment C, network meta-analyses can provide indirect evidence on the comparison between treatments A and C, which were not directly compared. However, due to the assumption that treatment B is comparable in the studies comparing A and B and those comparing B and C, indirect evidence from network meta analyses is especially prone to bias induced by conceptual heterogeneity (e.g., Mills, Thorlund, & Ioannidis, 2013).

Is the Intervention Effective for this Specific Patient?

Clinical research claims to produce results that are relevant to clinical practice. One type of evidence relevant to practice can be derived from the efficacy and effectiveness studies already described. Using the strategies described above (RCTs, naturalistic studies, meta-analyses), the

average effectiveness of an intervention can be evaluated (for specific patient groups). Intervention strategies that have been shown to be effective in studies with these designs are called "evidence based." This label indicates that these intervention strategies are promising for the treatment of patients with the respective symptoms and disorders. For a detailed description of research-supported treatments and their evidence base we refer to the website of the Division 12 of the APA (http://www.div12.org/psychological-treatments/, retrieved April 3, 2017). However, results from these studies are based on univariate or multivariate mean comparisons averaged over all patients. Interindividual variation is neglected and considered *error variance*. From this "average change," it can only be concluded that an intervention is effective "on average" (i.e. for the average patient), but not whether it is effective for a specific patient. While the research designs described so far provide evidence on a group level, the following designs are more concerned with evidence on an individual level.

Single Case Research

In single case studies, one specific area is investigated intensively. Due to the lack of controlled conditions, traditional case studies are limited with regard to generalizable conclusions. However, they are a source of generating new hypotheses and developing new intervention techniques. It is also possible to study rare events with single case studies. Having a high heuristic validity, case studies are often used when introducing new approaches and techniques in clinical psychology. Problems lie in the ambiguity of possible alternative explanations as well as in the questionable generalizability to other patients or situations. Part of these limitations can be addressed via single-case experiments or single-case quasi experiments. The basic principles of the single-case experiment are identical to those of group experiments. However, instead of comparing groups, the conditions are realized within persons over time. That is, observations or assessments of behavior of the patient are made repeatedly over time in different conditions. For experimental single-case research, a series of designs were developed (compare e.g., Barlow, Nock, & Hersen, 2009; Kazdin, 2013). Two of these designs, the *A-B-A-B design* and the *multiple baseline design* are briefly described in the following. In the *A-B-A-B design,* after assessment of the baseline (A), the intervention (B) is implemented, followed by another assessment of the baseline (A). During the repeated baseline phase, the effect that was produced during the first intervention phase is regularly reduced. Therefore, an additional intervention phase (B) follows in order to control the effect.

Multiple baseline designs use repeated assessments of the baseline in different situations or for different problem behaviors. If a single case experiment can be conducted with more than one participant, it could be even useful to vary the number of baseline assessments. If, for all tested individuals, their problems improve after the intervention started, despite varying numbers of baseline assessments, it is likely that these improvements are caused by the intervention.

Single case studies vary in the assessment of baselines and the way the intervention is conducted. Experimental or quasi experimental single case studies are analyzed via graphical approaches, analyses of variance and time-series models.

Patient-focused Research and Quality Management

Traditionally, the introduction of new therapeutic strategies or treatment approaches is based on a clinical idea proposed by a researcher or clinician. At this stage, research is seldom part of the development process. Therefore, it is pivotal to accompany the introduction of new teatment concepts with rigorous research prior to a broad dissemination of this approach. All too often practitioners need to save themselves from prematurely jumping on the bandwagon of a newly developed paradigm. It is therefore important for clinicians to be able to read,

understand, and integrate empirical studies in a way that allows them to evaluate an intervention's evidence base. However, even if several original studies and meta-analyses have shown that an intervention is effective on average, we cannot conclude weather this intervention works for a particular patient (e.g., Howard et al., 1996). A continuous monitoring of patient progress by means of repeated assessments of outcomes is therefore needed in any treatment. This kind of progress monitoring allows the information to be fed back to the therapists who are thus enabled to directly integrate the so gained knowledge to optimize their strategy (e.g., Lambert, 2007). This research design involves fine-meshed repeated assessments over the course of treatment. Due to its focus on the individual patient, this paradigm is called *patient-focused research* (Howard et al., 1996, Lutz, de Jong, & Rubel, 2015). Based on large-scale data from patients who have already been treated, it is possible to deduce predictions for the treatment course of a newly incoming patient. These predictions could support the selection of the most promising treatment for this new patient (e.g., DeRubeis et al., 2014; Lutz, Leon, Martinovich, & Stiles, 2007). That is to say, this patient could have a more positive prognosis if receiving treatment A than receiving treatment B. In that way *patient-focused research* is for psychological treatments, what *personalized medicine* or *precision medicine* is for medical treatments (Hamburg & Collins, 2010; National Research Council (U.S.) Committee on a Framework for Developing a New Taxonomy of Disease, 2011).

During the course of a treatment, these individual predictions could serve as benchmarks to which the actual course of change of a patient can be compared. As such, patient-oriented research provides evidence, which is *practice-based* and directly applicable to ongoing treatment (Castonguay, Barkham, Lutz, & McAleavy, 2013). This application of *patient-focused* research makes it a form of an ongoing *quality management* as well. Patient progress is compared repeatedly with what could be expected for that individual patient. If a patient is not progressing as intended, an adaption of the intervention may be necessary.

How, for Whom, and under which Conditions do Clinical Interventions Work?

Knowledge about the ingredients that make clinical interventions effective and the mechanisms through which they work can help to optimize our treatments, making them more effective and efficient. An important distinction to make at this juncture is that of *common and specific factors* of psychotherapeutic interventions. While common factors are those that are shared by all psychotherapeutic interventions, such as the expectancy of the patient that somebody will help him, the belief of the therapist that he provides an effective treatment and a trustful therapeutic relationship, specific factors are techniques that are unique to different therapeutic orientations. For example, the dispute of dysfunctional beliefs via Socratic questioning is a method that is predominantly applied in cognitive therapies, while the interpretation of dreams is a technique specifically present in psychoanalysis. There is an ongoing debate, whether *common* or *specific factors* are responsible for therapeutic change (e.g., Hofmann & Barlow, 2014). Process research investigates these mechanisms of action underlying clinical interventions.

Methods of Process Research

An extensive collection of data is common to all approaches of process research (e.g., through session-by-session assessments with questionnaires or video analyses). Therefore, from a technical point of view, process research profited strongly from the advancements of computer systems (e.g., touch screen data entries), which enable large datasets to be depicted and processed with respect to different aspects of the therapeutic process. The continuous application

of session report questionnaires at the beginning or end of each therapy session can, for example, document the individual progress of an intervention. By adding post hoc ratings of videotaped sessions, it can be determined which central mechanisms were realized at which point in time, and how they were related to treatment outcome. A central aspect is the selection of the appropriate observation entity. Single words, gestures, episodes, entire sessions, time intervals of different duration or entire treatment phases up to a number of sessions can be investigated (Crits-Christoph et al., 2013). In process research, *qualitative methods* are also often applied. These are more concerned with the qualities of phenomena than their quantification. By doing so, qualitative methods emphasize the meaning for the participants and concentrate on language use during the intervention (e.g., Lutz & Knox, 2014). Consequently, the analyzed data are primarily words, from which interpretations, constructs, and theories are deduced, which stipulate future qualitative and quantitative investigations (Kazdin, 2013).

Dismantling and Additive Designs

Studies that specifically test particular therapeutic ingredients are called dismantling studies. In these studies, intervention programs are dismantled and versions in which systematically specific components are left out are tested against each other in RCTs. If the effects of a program are reduced, if a specific ingredient is left out, this provides evidence for the importance of that specific component. In additive designs, an existing approach is tested against the same treatment plus a specific component that is newly added. In an RCT, whether the extended treatment is able to augment the effects of the traditional approach is then tested. Such research strategies enable the identification of potential mediators in the relation between the intervention (independent Variable; IV) and the outcome (dependent Variable; DV).

Mediators and Moderators of Clinical Interventions

If we are interested in the mechanisms through which treatments work, statistically this is a question of mediation. Mediation describes a specific relationship between three or more variables. In the simplest case, the relation between two variables, an independent variable (IV; e.g., treatment: yes or no) and a dependent variable (DV; e.g., depressive symptoms) comes about due to a mediator variable (MED; e.g., dysfunctional beliefs). If complete mediation is present, the total effect of the IV on the DV is mediated through the MED. That is to say, without a change in the MED (e.g., dysfunctional beliefs), there would be no difference in the DV (e.g., depressive symptoms) regardless of the IV (e.g., treatment or no treatment). However, that is a rare scenario in behavioral research. More often partial mediation is observed. In partial mediation the total effect does not go through the mediating variable. Rather the total effect splits into the indirect effect (IV $\rightarrow$ MED $\rightarrow$ DV) and the direct effect (IV $\rightarrow$ DV). Mediation is one way to identify the working mechanisms of clinical interventions. However, additional conditions must be met to establish a mediation as a causal mechanism. For an in-depth discussion of causality and mediation, refer to MacKinnon, Lockhart, Baraldi, and Gelfand (2013).

As in mediation analyses, moderator models also describe relations of three or more variables. However, in moderation analyses, the IV does not influence the moderating variable (MOD; e.g., sex). Rather, the MOD influences the association between the IV and the DV. If a treatment is more effective for woman than for men, sex is a moderator of the treatment effect. Thus, moderators tell us something about the differential effects of interventions (i.e. "for whom"). It is important to note that the difference between moderation and mediation is not always clear cut, and many combinations of both are possible (e.g., MacKinnon et al., 2013).

Summary

The present chapter provided an overview of the most important methods of clinical psychology. Clinical researchers as well as clinicians must be acquainted with these concepts in order to be able to advance clinical science and provide treatments with a firm evidence base. In times when clinical psychology is becoming more and more heterogeneous in terms of therapeutic approaches, clinicians must be able to use the existing evidence to separate the wheat from the chaff. Without a firm understanding of research methods, clinicians could hardly accomplish this task. However, we showed that knowledge on the evidence base of an intervention is not enough to succeed in everyday clinical practice. Rather, clinicians need methods to evaluate their own work and tailor treatments to the specific needs of their patients. On the road to improvement, it is critical to know where one's strengths and weaknesses lie. A continuous evaluation of personal clinical practice can thus help clinicians to improve their strategy for individual patients, as well as their general clinical abilities.

References

Barlow, D. H., Nock, M., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior for change* (3rd ed.). Boston: Pearson/Allyn & Bacon.

Castonguay, L., Barkham, M., Lutz, W., & McAleavey, A. (2013). Practice-oriented research: Approaches and applications. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (6th ed., pp. 85–133). John Wiley & Sons.

Crits-Christoph, P., Connolly Gibbons, M. B., & Mukherjee, D. (2013). Psychotherapy process-outcome research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 298–340). New York, NY: John Wiley & Sons, Inc.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PloS one*, *9*(1), e83875.

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine*, *363*(4), 301–304.

Hofmann, S. G., & Barlow, D. H. (2014). Evidence-based psychological interventions and the common factors approach: The beginnings of a rapprochement? *Psychotherapy*, *51*(4), 510–513.

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist*, *51*, 1059–1064.

Hunter, J. E., & Schmidt, F. E. (2004). *Methods of meta-analysis. Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Kazdin, A. E. (2013). *Research design in clinical psychology* (4th ed.). Harlow: Pearson Education Limited.

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, *17*(1), 1–14.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in medicine*, *21*(16), 2313–2324.

Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research*, *12*(3), 251–272.

Lutz, W., de Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research*, *25*(6), 625–632.

Lutz, W., & Knox, S. (2014). Quantitative and qualitative methods for psychotherapy research: introduction. In W. Lutz & S. Knox (Eds.), *Quantitative and Qualitative Methods in Psycho-therapy Research* (pp. 1–5). New York, NY: Routledge.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology*, *54*, 32–39.

MacKinnon, D. P., Lockhart, G., Baraldi, A. N., & Gelfand, L. A. (2013). Evaluating Treatment Mediators and Moderators. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 262–286). New York, NY: Oxford University Press.

Mills, E. J., Thorlund, K., & Ioannidis, J. P. (2013). Demystifying trial networks and network meta-analysis. *BMJ*, *346*, f2914.

National Research Council (U.S.) Committee on A Framework for Developing a New Taxonomy of Disease. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press (U.S.).

Ogles, B. M. (2013). Measuring Change in Psychotherapy Research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th ed., pp. 134–166). New York, NY: John Wiley & Sons, Inc.

Rockett, I. R. (1999). *Population and health: An introduction to epidemiology* (2nd ed.). Washington, DC: Population Reference Bureau.

Rounsaville, B. J., Carol, K. M., & Onken, L. S. (2001). A stage model of behavioral therapies research: getting started and moving on from stage I. *Clinical Psychology: Science and Practice*, *8*, 133–142. doi:10.1093/clipsy.8.2.133.