



HAL
open science

Evaluating complex interventions: Perspectives and issues for health behaviour change interventions

Cyril Tarquinio, Joël Kivits, Laetitia Minary, Joël Coste, François Alla

► To cite this version:

Cyril Tarquinio, Joël Kivits, Laetitia Minary, Joël Coste, François Alla. Evaluating complex interventions: Perspectives and issues for health behaviour change interventions. *Psychology and Health*, 2015, 30 (1), pp.35-51. 10.1080/08870446.2014.953530 . hal-01573571

HAL Id: hal-01573571

<https://hal.univ-lorraine.fr/hal-01573571v1>

Submitted on 9 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Evaluating complex interventions: Perspectives and issues for health behaviour change interventions

Cyril Tarquinio, Joëlle Kivits, Laetitia Minary, Joël Coste & François Alla

To cite this article: Cyril Tarquinio, Joëlle Kivits, Laetitia Minary, Joël Coste & François Alla (2015) Evaluating complex interventions: Perspectives and issues for health behaviour change interventions, *Psychology & Health*, 30:1, 35-51, DOI: [10.1080/08870446.2014.953530](https://doi.org/10.1080/08870446.2014.953530)

To link to this article: <http://dx.doi.org/10.1080/08870446.2014.953530>



Accepted author version posted online: 20 Aug 2014.
Published online: 21 Oct 2014.



[Submit your article to this journal](#)



Article views: 1012



[View related articles](#)



[View Crossmark data](#)



Citing articles: 11 [View citing articles](#)

Evaluating complex interventions: Perspectives and issues for health behaviour change interventions

Cyril Tarquinio^{a,b}, Joëlle Kivits^{b,c}, Laetitia Minary^{b,d,e}, Joël Coste^{b,f} and François Alla^{b,c,d,e*}

^aLaboratoire de psychologie – Université de Metz, METZ cedex 1, France; ^bUniversité de Lorraine, Université Paris Descartes, Apemac, EA 4360, Nancy, F-54 000, France; ^cUniversité de Lorraine, Faculté de Médecine, Ecole de Santé Publique, Nancy, F-54 000, France; ^dCHU Nancy, Epidémiologie et Evaluation Cliniques, Nancy, F-54 000, France; ^eINSERM, CIC-EC, CIE6, Nancy, F-54 000, France; ^fDépartement de Biostatistique, Pavillon Saint-Jacques, Hôpital COCHIN, 27 rue du Faubourg Saint-Jacques, 75674, Paris Cedex 14, France

(Received 19 August 2013; accepted 6 June 2014)

Objective: Health behaviour change interventions (HBCIs), used in health education, health promotion, patient education and psychotherapy areas, are considered complex interventions. The objective of this article is to discuss the value and limitations of using randomised clinical trials (RCTs) to assess HBCIs.

Methods: A scoping review of the literature was conducted to identify the main challenges of using RCTs for evaluating HBCIs. The issues were illustrated by case studies selected from research conducted by our multidisciplinary team.

Results: In complex interventions, effects are produced not only by the intervention, but are strongly linked to context. Issues relating to transferability of results are therefore critical, and require adjustments to the RCT model. Sampling bias, biases related to the experimental conditions and biases due to the absence of double-blindness were examined and illustrated by case studies.

Conclusion: The results underline the importance of a multidisciplinary approach. They call for adapted or alternative evaluation models that overcome the limitations of RCTs.

Keywords: behaviour; psychology; public health; health promotion; health education; randomised controlled trial; evaluation studies; bias; complex interventions

1. Introduction

The areas of health education, health promotion, patient education and clinical health-care, including psychotherapies (Campbell et al., 2000) are among the most active in health behaviour change. Thus, they have fostered, launched and developed health behaviour change interventions (HBCIs). These interventions could be considered ‘complex interventions’ (Craig & Pettecrew, 2013), which are constructed using several interacting components (Craig & Pettecrew, 2013). These components include behaviours, behaviour parameters and methods of organising those behaviours, and they may have

*Corresponding author. Email: francois.alla@univ-lorraine.fr

an effect at the individual, service and/or population level (Craig & Petticrew, 2013). Furthermore, they can have multiple targets (individuals, groups, communities and organisations) and their outcomes are numerous and variable. Their effects are thus the product not only of the interventions themselves, but also of the contexts in which they unfold (Cambon, Minary, Ridde, & Alla, 2012).

These interventions all face the same challenge that of demonstrating efficacy and responding to the evidence-based exigencies of health sciences. While they remain the gold standard in the biomedical field, randomised clinical trials (RCT) are difficult to adapt to the inherent complexity of HBCIs. Complex interventions may require a less linear and more flexible evaluation model (Craig & Petticrew, 2013). The usual distinction between efficacy and effectiveness is, in the case of complex interventions, an important aspect to consider. The efficacy of an intervention is classically defined as its effect under ‘ideal conditions’, while effectiveness is its effect under normal conditions. However, for complex interventions, this binary distinction could be not relevant, particularly as the ‘idealized condition’ can hardly be created. It is also irrelevant from a generalisability and transferability perspective as the environment and conditions of implementation are themselves determinants of the intervention outcomes. Some authors have thus put forward more nuanced definitions, e.g. the typology of evaluation studies proposed by Victora, Habicht, and Bryce (2004). This is not without consequences, since such a position has led numerous researchers to work on psychological treatments comparable to that of the drug and to pay almost exclusive attention to the outcomes of brief, focused therapies and their overgeneralisation to very different patient groups. Let us recall, in this regard, that in 1996, the National Institute of Mental Health (NIMH) launched an appeal for more research aimed specifically at real effectiveness (Krupnick et al., 1996). Indeed, in many cases, controlled randomised trials focused on experimental populations (e.g. recruited among university students) and not in common populations treated in clinical settings.

The purpose of this article is to discuss the value and limitations of using RCTs to assess HBCIs. This questioning transcends all disciplines that address the evaluation of behaviour change. We will show why such approaches may be inadequate, whether in public health or psychology. More specifically, we will discuss the question of evaluating psychotherapies. Finally, we will consider biases that highlight the challenge of evaluating HBCIs.

2. Methods

A scoping review was conducted to identify the main challenges of using RCTs for evaluating HBCIs and to describe methodological adjustments or alternatives discussed and/or tested by researchers. Scoping studies can be used for rapid mapping of the key concepts underpinning a research area as well as the main sources and types of evidence available. They can be undertaken as stand-alone projects in their own right, especially when the area is complex or has not yet been reviewed comprehensively (Arksey & O’Malley, 2005; Mays, Roberts, & Popay, 2001). In this case, methodological papers, research reports, reviews and opinion papers from scientific databases and elsewhere (reports and books), were identified in a stepwise approach.

Case studies stemming from past or current research conducted by our multidisciplinary team were selected as being illustrative of challenges in RCTs. The implications of these case studies concerning RCT methodology are discussed in the results section.

3. Results

3.1. Evidence in HBCIs: nature and challenges

Discussing the interests and limitations of RCT for health intervention evaluation inevitably raises the question of experimentation and the nature of the evidence that is produced (WHO, 2001). Experimental design defines a classical process to explain the results of an intervention by neutralising the influence of contextual factors by balancing their effects between groups. It constrains the methodological aspects of the intervention. RCT is recognised in the health and medical sectors as a gold standard for demonstrating the causal link between an intervention and an effect (Rychetnik, Frommer, Hawe, & Shiell, 2002). A clinical trial is indeed considered the method with the strongest internal validity, as well as the strongest scientific inference, minimising the risk of bias. Given that there is an inverse strength relation between internal and external validity (Campbell & Stanley, 1963), the ascendancy of clinical trial over any other design is based on the premise of the superiority of internal validity over external validity. In other words, RCTs are used not so much because they represent ‘real situations’, but rather because experimental conditions, ‘protected’ from reality, are better able to generate definitive evidence. While justified in simple contexts such as drug trials, this epistemological position is disrupted and difficult to sustain when the object of evaluation is a complex intervention including health behaviour change. Effectively, the neutralisation of contingent factors constitutes both the underlying strength of experimental design – scientific rigour being a matter of controlling elements – and its main limitation – as it does not aim to explain real-life effects. A real situation – a patient with his/her psychotherapist, a patient education session or a health education programme – becomes a possible scenario derived from the experimental model. This calls into question the relevance of demonstrated causality: the limitation of experimental design is precisely that it minimises contextual factors that may, in fact, be essential to the success of the intervention.

Moreover, depending on the nature of the intervention, it is sometimes not possible, for practical or ethical reasons, to carry out an RCT. Therefore, the evidence-based approach may not favour the most effective interventions, but those that are easiest to implement in the context of a trial. This is an important limitation for the development of evidence-based health promotion (Rychetnik et al., 2002).

There is thus much current debate concerning the relevance of the conditions under which evaluation must be conducted.

3.1.1. Challenges to RCTs in psychotherapy

Psychotherapies mobilise complex and multiple processes. They are therefore among the most difficult areas to evaluate. There is a wide gap between the multiple components that constitute the psychotherapist’s intervention and the components that can actually be evaluated. It is therefore necessary to make choices. In so doing it is important to ensure that this reduction does not lead to invalidation of the actual scope of the results. From a clinical perspective, initial diagnosis goes beyond the concept of symptom or illness and takes account of the problem of the person in her/his different components: developmental, historical, current – including context. However, evaluation for

research purposes cannot take account of all of these parameters, which may lead to over simplification of the intervention.

Recent research by Mynors-Wallis, Gath, Day, and Baker (2000) illustrate this problem: it leads to the surprising result that severe depression can be treated in six sessions with an expression and problem-solving technique led by a doctor or nurse who have been quickly trained in the technique. Similarly, the attempt to identify psychological treatments or effective interventions whose effects are equal to those of a drug has resulted in almost exclusive concentration on the results of short, focused therapies and on their inappropriate transposition to very diverse groups of patients. This avoids reflection on the actual objective of the research and on the fact that the results obtained are actually far removed from real-life experience (Chambless & Ollendick, 2001).

Analyses of the components that could influence the outcome of psychotherapy have shown that the main psychotherapies obtained similar results. This has drawn attention to the importance of 'common factors' in the different psychotherapies, particularly relationship psychotherapy, which is directly linked to the therapist as an individual (Wampold & Brown, 2005) who constitutes an important predictive element for the results. Thus, the interventions of therapists showing empathy, human warmth, compassion and sensitivity are more conclusive (Green & Herget, 1991). In this respect, the research tends to show that a positive therapeutic alliance between patient and therapist is the best indicator of results in psychotherapy. The therapeutic alliance has emerged as an important variable in psychotherapy process/change for various schools of psychotherapy (Elvins & Green, 2008). Meta-analysis of studies in both adult (Martin, Garske, & Davis, 2000) and child (Shirk & Karver, 2003) mental health research have shown a consistent association between measures of therapeutic alliance and treatment outcome, particularly when alliance is measured at early stages of treatment (Howard, Turner, Olkin, & Mohr, 2006). These studies, as well as findings within trials (Klein et al., 2003), have also indicated the consistency of the alliance effect across different types of treatment. Such evidence is consistent with clinicians' views (Kazdin, 2007) and patients' views (Johansson & Eklund, 2003) describing the therapeutic relationship as a key determinant of treatment success. What can seem to be a very general ingredient within reach of everyone with a capacity for empathy and some charisma (Strupp & Hadley, 1979) has been shown by analysis to be an extremely subtle construction which depends on the practitioner's competence, particularly in terms of the degree of adaptability to defence mechanisms. The results of Barber et al.'s study (Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000) suggest that although early in treatment the alliance might be influenced by prior symptomatic improvement, it remains a significant predictor of further improvement even when prior change in depression is partialled out. The role of the alliance as a potential causal factor in patient improvement is consistent with the theoretical and therapeutic role suggested by various psychodynamic and experiential theorists (Elvins & Green, 2008). Of all the variables studied, the therapeutic alliance has proved to be the best predictor (Krupnick et al., 1996). The correlation between level of alliance and symptomatic evolution of the patient is of the same order for all four groups (both psychotherapies, drug therapy and placebo group). Finally, the involvement of the alliance in outcome variance is much higher (about 21% of explained variance) than the therapeutic technique itself (only 2% of explained variance). But the alliance is not an intervention in and of itself.

Its emergence reflects, rather, the appropriate use of other elements of the therapeutic process (Stiles, 1999). Examples of direct clinical training in therapeutic alliance have not, for example, yielded particularly positive results (Horvath, 2005). Indeed they have even shown deterioration in outcomes following training (Henry, Schacht, Strupp, Butler, & Binder, 1993). It is not, therefore, possible to focus exclusively on relational aspects to the detriment of technical aspects, or on common factors as opposed to specific agents. Understanding what happens in psychotherapy, as recommended by the current research phase, requires models that integrate these two dimensions in a model of understanding that is both more global and better contextualised.

Another common factor is the patient's motivation, but in fact the literature seems to show that each factor only truly intervenes in interaction with the others (Valbak, 2004). Blatt, Sanislow III, Zuroff, and Pilkonis (1996), for example, have investigated the influence of the quality of the therapist on the results. Their study included 28 highly experienced and well-trained therapists whose adherence and competence were controlled. Nevertheless, a calculation of the mean effectiveness of each therapist revealed three different groups, i.e. those who were: (1) not very effective, (2) relatively effective and (3) more effective. However, the group divisions did not correlate to types of treatment. In a study by Kim, Wampold, and Bolt (2006), about 8% of the variance of the results was attributable to the quality of the therapist, while 0% was attributable to the type of treatment.

3.2. *Specific or emphasised bias in HBCIs*

Because HBCI's effects involve potential interactions between the intervention and its environment (such as individual factors), some bias encountered in clinical research can be reinforced in the context of HBCI evaluation. Biases that can occur during RCTs, e.g. selection, performance, attrition and detection biases are well documented, especially for clinical research (Higgins & Altman, 2008). Our aim, in the area of HBCIs, is to focus more specifically on biases inherent in the experimental method, both in the method itself (e.g. standardisation of the intervention) and in any adaptation made to take account of constraints related to the subject (e.g. the impossibility of blinding). The biases described below are for the most part not specific to the field of HBCI. However, what is only a limitation in clinical research becomes a major obstacle in the field of HBCI, an obstacle that not only affects the transposability and dissemination of results, but also calls into question the actual nature of the evidence obtained. We have classified these biases into three groups, while recognising that their determinants and effects are interrelated.

3.2.1. *Sampling bias*

The characteristics of subjects eligible for recruitment in trials are different from those of other subjects; for methodological reasons (e.g. reduce the variance and thus the number of subjects needed; reduce the risk of loss to follow-up, limit the risk of interactions with other treatments), trials include more homogeneous, younger patients, without co-morbidities, etc. This sampling bias (Porta, 2008), has an impact on the external validity of the results. It is particularly important to consider this bias when the

causal effect of an intervention can vary across individuals (Weisberg, Hayden, & Pontes, 2009). In particular, the beneficiaries' actual need with respect to the intervention must be taken into consideration (Cambon et al., 2012). This concept, which is based on the theory that the same dose will have less effect if there is less need for it, has been defined as a 'curvilinear dose–response association' (Victora et al., 2004).

In the field of psychology, this simplification of clinical situations may affect the very validity of the results. Research shows weak clinical validity with regard to patient recruitment and relevance of diagnosis, often limited to one symptom. This is especially true since the advent of DSM (Diagnostic and Statistical Manual of Mental Disorders), which remains a perfect standard for efficient and simple categorisation of a population, without, however, the ecological dimensions of this population and the complexity of the disorder being thereby limited. This is an important point because, in psychology, disorders only have pathological value in their diagnostic environment. An isolated disorder, such as a social phobia, may have a totally different significance if it emerges alone (discomfort in groups) rather than as part of a psychotic disorder.

Westen and Weinberger (2004) pointed out the fact that if a patient with one symptom responds to treatment, it does not mean that another patient with a group of associated symptoms or diseases, which is the most frequent case in real clinical situations, would respond similarly to the same treatment. This has been criticised by Guthrie (2000), who observed that

this leads to a rather curious situation in which most of the studies on psychological treatment target the patients who, in clinical practice, would be treated with drugs, and exclude the patients to whom, in clinical practice, one would propose a psychological treatment.

Finally, Ablon and Jones (2002) summarise the situation succinctly, by stating that randomised controlled trials test a somewhat artificial treatment in an artificially controlled configuration with atypical patients, so they have limited capacity for generalisation to the real world of mental health care delivery.

In addition to the selection by recruitment process, self-selection – the subject choosing to participate or not in a study – could reduce the validity of the trial if determinants of choice are also determinants of outcome. This bias is called self-selection bias or volunteer bias (Heiman, 2002). Beyond the effect of randomisation itself, the experimental condition itself (compared to an observational design) can be problematic. In fact, accepting and being motivated to participate in a study, or to be observed, can involve individual factors like empathy, level of education, personality, knowledge, opinions, health behaviour or interest in the topic of research (Colsher & Wallace, 1989; Heiman, 2002; Musick, Wilson, & Bynum, 2000; Oswald, 2003). These factors constitute a major issue, especially as they are generally not routinely measured and will thus not be taken into account during dissemination. For example, the TABADO study (Minary et al., 2009, 2013) aimed to evaluate the efficacy of offering a community-based intervention, a smoking cessation programme, as part of a comprehensive approach to prevention in a population of young trainees in VTCs. We performed psychological measures in young people who were potentially eligible to join the

TABADO programme before the programme was offered to them. Volunteers participating in the programme (25% of the population) were psychologically different from non-volunteers. According to the Brief Cope scale (Muller & Spitz, 2003) and the Self-Regulation Skills Scale (Husiman, Maes, De Gucht, Chatrou, & Haak, 2009), they were more susceptible to self-criticism (6.6 vs. 5.4, $p = .04$), seeking social support (10.5 vs. 9.1, $p = .012$) and instrumental social support (4.1 vs. 3.5, $p = .032$), and were predisposed to active coping (facing their problem directly and openly; 3.7 vs. 3.3, $p = .043$) in stressful situations. However, active coping is related to smoking cessation, the major endpoint of the study (adjusted OR = 1.74; IC = [1.21–2.50]; $p = .003$). This example demonstrates that volunteers for participation in a HBCI might have specific psychological characteristics, a result that underlines a risk of volunteer bias in such interventions.

In open trials, such as behavioural change intervention trials, the self-selection process also arises after randomisation, leading to both internal and external validity issues. This is similar in randomised trials using Zelen's method (Zelen, 1979) and in cluster randomised trials. This selection leads to a confounding factors imbalance between intervention and control groups (Kannus et al., 2000; Torgerson & Torgerson, 2008). We can illustrate this type of bias through the Regul'aps study (Tessier, 2008). In French primary schools, children participate in three compulsory hours of physical education (PE) each week. The objective of Regul'aps was to evaluate whether splitting up the three hours into three or four sessions (vs. one or two sessions) of PE per week had an effect on speed of BMI increase and on health-related quality of life (HRQoL) over the school year. Methodological difficulties arose when implementing the RCT. When presented to teachers, the trial generated three types of reaction: refusal, acceptance and acceptance under the condition that they would choose the group to be studied (considering the intervention would affect their timetable). The question was whether to include the classes whose teachers had chosen the intervention without going through randomisation within the group of classes accepting randomisation. In Regul'aps, these teachers' classes finally constituted a third group, the randomisation refusal group, in addition to the intervention and control groups. A further objective was to decipher whether the groups were differentiated in terms of parental refusal rate, age and BMI (Tessier, Vuillemin, & Briançon, 2006). A comparative analysis between the randomisation refusal group and the randomisation accepting group led to the result that the ages were similar but the individuals differed at baseline in terms of BMI – the principal endpoint of the study (17.1, ± 2.8 vs 16.7, ± 2.8 , $p = .023$) – and parental refusal rate (15.2% vs. 9.8%). This study therefore illustrates sampling bias. The teachers' willingness to participate in the RCT was clearly dependent on several parameters such as their rejection of the new/supplementary project, their knowledge of available material resources and the children's social environment.

In summary, sampling bias involving a difference of characteristics between participants in the intervention and the – participants, has to be considered especially in the context of public health interventions: on the one hand, for a same intervention dose, the effect may vary according to the beneficiaries needs; on the other hand, the factors inducing the subject's participation in a study – such as psychological traits – may also be factors impacting the intervention outcome.

3.2.2. Biases related to experimental conditions

Experimental conditions differ from those of real life in many significant ways: volunteer and trained professionals; standardised procedure; specific context (e.g. specialised research units); and specific measurements of behaviour. This experimental context, with a specific process involving a particular relationship between health professional and patient, not only selects particular individuals but also involves particular behaviours. A known example is the Hawthorne effect, defined as an improvement in people's performance due to their effort and motivation when they receive an intervention and when they know they are being evaluated (Fernald, Coombs, DeAlleaume, West, & Parnes, 2012; Mayo, 1933).

This raises the issue of the transferability of results. Transferability has been defined as the extent to which the measured efficacy of an intervention could be achieved in another setting (Wang, Moss, & Hiller, 2006). It must be distinguished from the concept of applicability, defined as the extent to which an intervention process could be implemented in another setting (Wang et al., 2006). Many factors – related to experimental conditions – could influence the transferability of results of HBCIs (Cambon et al., 2012; Victora et al., 2004), either directly (outcomes are not transferable because the terms and conditions for implementing the intervention are different) or indirectly (for the same implementation modalities, different outcomes are obtained). These elements may include population characteristics, the environment in which the intervention takes place, the quality of the relationship between intervention participants, and the adaptations of the intervention made on the ground by the caregiver. In most real-life situations, the dose received by the population is likely to be smaller than in any type of trial (Victora et al., 2004). Thus, what use is evidence produced under experimental conditions if it cannot be used in real-life conditions? Indeed, for the intervention's stakeholder, the question is not, 'Does it work theoretically?' but rather, 'Does this work potentially in my population, considering my resources and implementation context?' (Trewick & Zwarenstein, 2009). Thus, in HBCIs, the challenge today is to be able to measure the effect of contextual elements to ensure that the studied intervention would produce the same results when implemented in another context.

Another aspect of HBCIs must be taken into account: any complex intervention necessarily adapts its evaluation model as it is accepted that results may vary according to practitioners' skills, experience, population motivation, etc. The principle of standardisation of experimental design – an individual receives the same intervention under the same conditions as any other person participating in the experimental design – is not sustainable. In psychotherapies or health education, one can easily see the impossibility of guaranteeing conditions of standardisation. Standardisation is opposed to the necessary adaptation of complex interventions to their implementation context. As reviewed by Datta and Petticrew, '[a] degree of flexibility in the design and implementation of interventions was advocated by a number of writers with the aim of ensuring that interventions could be adapted to both local circumstances and to patients' needs' (Craig & Petticrew, 2013). In the same way, Hawe et al. indicate that in many complex interventions, instead of standardising the form of a component (e.g. using exactly the same patient education materials), it may be more appropriate to standardise its function while tailoring the form to the context (e.g. allowing materials to be adapted to local cultural styles and literacy levels) (Hawe, Shiell, & Riley, 2004, 2008).

In the PRALIMAP school-based RCT (Bonsergent et al., 2013), the objective was to evaluate the effectiveness of three strategies aimed at preventing overweight and obesity among adolescents in a high school setting: education strategy, environment strategy, screening and care strategy. The two-year change in outcomes was more favourable in the 12 screening and care high schools compared to the non-screening ones (i.e. other strategies or non-strategy): a .11 lower increase in BMI ($p = .03$), a .04 greater decrease in BMI z-score ($p = .02$), and a 1.71% greater decrease in overweight/obesity prevalence ($p = .04$). Education and environment strategies were not more effective than no intervention. A recent study analysing the effects of the PRALIMAP intervention from the stakeholders' perspective actually showed that the 'turnkey' nature of PRALIMAP, imposed by the experimental design, was considered unacceptable by stakeholders and was potentially ill-suited to local needs and constraints (Trompette, Kivits, Minary, Cambon, & Alla, 2013).

PRALIMAP illustrates the limitations of standardisation when applied to HBCIs. Contextual elements, such as actors' perceptions of the trial, local constraints, actors' subjectivity in delivering an intervention, have an impact on the effects of intervention, particularly attitudes and behaviours, conscious or not, towards the intervention and/or the targeted population that may lead to systematic difference. The same issue arises in the evaluation of psychotherapy. Psychotherapy is extremely difficult to 'standardize', so its procedures and techniques can be used in the same way by all therapists without taking account of their training and personality. Each meeting between patient and therapist has certain unique characteristics, with the potential to produce 'something' which cannot be fully predicted or completely 'standardized'. A clinician is more than a technician; otherwise he would never be able to handle the complex problems that arise in clinical practice (Westen & Weinberger, 2004).

Thus, experimental conditions involve a context which is non-representative of real life and which may induce changes in behaviour by stakeholders and beneficiaries. In addition, the standardisation required by the experimental method cannot be guaranteed particularly in the context of intervention of behaviour change as it implies interactions between the intervention and the characteristics of each individual.

3.2.3. Biases due to the absence of double-blindness

The double-blind method is one of the major principles of randomised trials. Health behavioural change evaluation trials are generally by nature open trials. Biases in open trials have been well described (Schulz & Grimes, 2002). In the area of HBCIs, these biases could be reinforced.

The double-blind method avoids the subjectivity of the subject. Knowledge of the intervention received and perceptions of that intervention can affect participants' psychological or physical responses and behaviours. Participants who know they have been assigned to a group that will receive a new treatment might have specific expectations or apprehensions. Those assigned to standard treatment might feel deprived, or they may be more likely to use other forms of care (Schulz & Grimes, 2002). Where participants have a strong preference for one of the interventions but are allocated to the other, this can lead to the phenomenon of 'resentful demoralization'. (Brewin & Bradley, 1989; Torgerson & Torgerson, 2008). 'Resentful demoralization' is a well-recognised bias that may arise in control group participants, wherein the belief that they are not

receiving a desirable treatment negatively affects their attitude and behaviour and, as a consequence, the outcome results (Torgerson & Torgerson, 2008). If participants are not provided with the expected therapy, they may feel disappointed and have negative expectations of their ongoing involvement and these negative effects in the control group may in fact account for the apparently positive results of the experimental treatment. Given that trial study participants are typically highly vulnerable, it might be important to consider this effect in study design (Dunn, 2002). Moreover, an ineffective intervention administered in the context of RCTs may produce beneficial effects. This is the well-known placebo effect (Schulz & Grimes, 2002).

The double-blind method also neutralises the subjectivity of the investigator, who may otherwise adopt particular attitudes and behaviours, conscious or not, which may have a direct effect on the outcome or its measurement (e.g. detection bias) and which may lead to systematic differences in the care provided to the participants in groups other than the one receiving the intervention under investigation (i.e. performance bias) (The Cochrane Collaboration, 2004). The attitudes of an un-blind investigator for or against an intervention can be also transferred to participants (Schulz & Grimes, 2002).

Lastly, the double-blind method ensures objectivity in the measurement of judgment criteria. A recent review has shown that effect estimates were exaggerated in trials with subjective outcomes when there no blinding. In contrast, there was little evidence of bias in trials with objective outcomes (Wood et al., 2008). In the same way, lack of blinding has been linked to 30% exaggerated treatment effect estimates in a meta-analysis of studies with subjective outcomes (Wood et al., 2008).

In conclusion, by their nature, public health interventions do not allow the use of evaluative methods implying blind. Also, knowledge of the received intervention and the perception of that intervention can generate responses and psychological or physical behaviour on the part of beneficiaries and stakeholders which may bias its results.

4. Discussion

This article has presented and illustrated the main limitations of RCTs for the evaluation of HBCIs. Potential biases were examined and illustrated by case studies. We have shown how these biases, while they are for the most part not specific to the field of HBCIs, call into question the actual nature of the evidence obtained. For these complex interventions, effects are not only produced by the intervention, but are strongly linked to context. Issues relating to transferability of results are therefore critical, and require adjustments to the RCT model which might enable the limitations of such trials to be overcome. These issues have been explored for a long time, especially in the social and educational sciences (Campbell & Stanley, 1963), in which adaptations and alternative models to RCTs have been developed. These are now being disseminated in the health sector. Four basic types of methods are used for this purpose: (1) RCT adaptations; (2) process/mechanisms analysis and/or qualitative methods; (3) non-experimental methods; and/or (4) pilot study implementation.

4.1. Adaptations

To overcome the limitations of RCTs, some authors have proposed adaptations to RCTs to bring them closer to real-life conditions: e.g. pragmatic RCTs designed to test the

effectiveness of an intervention in routine clinical practice (Roland & Torgerson, 1998; Schwartz & Lellouch, 1967); the fully randomised preference design, developed to account for subject preference in open trials (Torgerson & Torgerson, 2008); cluster randomised trials, defined as experiments in which entire social units or clusters of individuals, rather than independent individuals, are randomly allocated to intervention groups (Donner & Klar, 2000).

These adaptations make it possible to test interventions in conditions closer to real life, i.e. to test the effectiveness of interventions and thereby maximise their transferability (Cambon et al., 2012).

4.2. Process/mechanisms analysis and qualitative methods

A second category of ‘adjustments’ to RCTs includes evaluation models that take into account the fact that there are certain aspects of HBCIs that cannot be grasped. They therefore focus on the operative mechanisms: these alternative models aim to understand why a specific result was obtained and what could have contributed to it. Adapted models incorporate the analysis of process, components and mechanisms into evaluation models. They involve use of qualitative or mixed research methods.

A first adjustment is that of process evaluation within RCTs (Oakley, Strange, Bonell, Allen, & Stephenson, 2006): the principle is not to disregard the trial, but rather to include an evaluation of the process within the experimental design, to understand the explanatory elements that may influence the outcome, such as ‘the implementation, receipt, and setting of an intervention’ (Oakley et al., 2006). This kind of evaluation responds to some of the challenges of complex intervention. It is, for example, suited to multisite trials, such as PRALIMAP (Legrand et al., 2012), as it is likely to highlight why a standard intervention may generate different results at different sites, thus contributing to improved transferability. Recently Bonell, Fletcher, Morton, Lorenc, and Moore (2012) have proposed the concept of ‘realistic randomized controlled trial’, to integrate the components of an intervention and their interactions into the evaluation. Realistic randomised controlled trials are developed as a model balancing experimental design with the necessary considerations as to which intervention works, for whom, under what circumstances and within what context. Process evaluation trials and realistic RCTs mark an advance in complex intervention evaluation. The complexity of HBCI is taken into account. Attention is paid to parameters that cannot be standardised or controlled, such as individuals’ perceptions. These models contribute to the methodological and conceptual understanding of intervention in the health area. Along with cluster RCTs, process evaluation trials and realistic RCTs aim to improve theorisation about the relationship between social phenomena and health (Fuller & Potvin, 2012). These models nevertheless retain the classic RCT as the standard, a fact that is currently under discussion, notably by realistic evaluators (Marchal et al., 2013). It still seems important to deepen and improve understanding of the mechanisms at the heart of HBCI.

The above adapted evaluation models all use qualitative and mixed methods, combining qualitative and quantitative research. Even before identifying the need for alternative evaluation designs, the literature calls for the uninhibited use of qualitative research methods in the field of public health (Jack, 2006; Meyer, 2000; Ridde & Mak, 2009). Indeed, quantitative methods present limitations for understanding components, processes and mechanisms, and for understanding individuals and groups’ perceptions,

moving contexts, indirect causal effects. The ‘how’ question cannot be satisfied if numbers are the only rule. Researchers have now addressed the usual criticisms of qualitative research methods. They are not ‘easy’ methods – they require time and strong methodological skills and are based on rigorous validity rules – yet they are different from quantitative methods. They also produce useful data; used in an evaluation context, they contribute to improving and correcting health interventions (Ridde & Mak, 2009).

Recently, in the evaluation field, the combined use of quantitative and qualitative methods has been conceptualised and defined as mixed methods research (Guevel, Pommier, & Jourdan, 2013). While relatively new in the field of health, mixed methods research approaches are well known in the field of education, where their use is accepted within the realistic paradigm (Blaise, Marchal, Lefèvre, & Kegels, 2010; Ridde & Haddad, 2013; Ridde, Robert, Guichard, Blaise, & Van Olmen, 2012).

4.3. *Non-experimental or quasi-experimental design*

We saw previously that RCTs, considered the gold standard in evidence-based medicine, are not always appropriate for evaluating complex interventions. Thus, we are witnessing a paradigm shift in which observational studies (or natural experiments Petticrew et al., 2005) and quasi-experimental studies (where random assignment is lacking) are being considered based on their capacity to produce data under ‘real-life’ conditions (Campbell & Stanley, 1963). The growing interest in comparative effectiveness research (CER) has led to a new interest in quasi-experimental and non-experimental studies due to their greater external validity (Department of Health and Human Services [HHS], 2009).

4.4. *Pilot study implementation*

Beyond evaluative methodology, a pilot study would help the investigator examine the key uncertainties identified during the development of an intervention (Craig & Petticrew, 2013). Thus, four groups of reasons for implementing a pilot study were proposed by Thabane et al. (2010): Process; Resources; Management; Scientific. Such a study should incorporate formative and process evaluation that would prevent some of the pitfalls in intervention delivery. Thus, the intervention content and the delivery could be investigated carefully according to the context of the intervention (Power et al., 2004) so as to maximise the likelihood of delivering a definitive trial that is high in terms of both internal validity and external validity (Bugge et al., 2013).

5. Conclusion

In conclusion, the question is not whether RCTs must be promoted. Of course, they must. The problem is much broader, as such a methodological choice is not without consequences, both on the nature of the data collected and on their significance, as well as on their appropriate interpretation. RCTs can answer the question of efficacy, all other factors being equal. This obvious fact is very often neglected, even when it fundamentally relativises the heuristic potential of the results obtained, which de facto must be considered not only partial, but also prejudicial, as they are the product of an arbitrary and debatable construction of what the researcher considers to be a variation of

reality – a reality devoid of complexity. We intuitively understand this same complexity when it is expressed in real life, but we are fundamentally unable to comprehend it methodologically. Evaluation must remain prominent in the field of health. It is therefore essential to have a methodological structure capable of equipping researchers in this field. The health context includes research and its operation in a particular ethical dimension, which requires the researcher to ask him/herself about the methods that he/she uses to generate knowledge and guidelines. This inevitably involves an epistemological questioning of approaches. This is what this text has tried to contribute, and is why there needs to be a critical, yet constructive, look at what might be more ecological devices capable of capturing a finer and more complex reality than efficacy alone. We can hope that such an investigation will result from a rapprochement between disciplines like public health, sociology and health psychology. While the objective is the same the method of attaining it requires major epistemological questioning which will doubtless enrich the way we do research.

References

- Ablon, J. S., & Jones, E. E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH treatment of depression collaborative research program. *American Journal of Psychiatry*, *159*, 775–783.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*, 19–32.
- Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology*, *68*, 1027–1032.
- Blaise, P., Marchal, B., Lefèvre, P., & Kegels, G. (2010). *Au-delà des méthodes expérimentales: l'approche réaliste en évaluation. Réduire les inégalités sociales en santé* [Beyond the experimental methods: The realistic evaluation approach]. In L. Potvin, M. J. Moquet, & C. Jones (Eds.), *Reducing social inequalities in health*. Paris: INPES.
- Blatt, S. J., Sanislow, C. A., III, Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: Further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, *64*, 1276–1284.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science and Medicine*, *75*, 2299–2306.
- Bonsargent, E., Agrinier, N., Thilly, N., Tessier, S., Legrand, K., Lecomte, E., ... Briançon, S. (2013). Overweight and obesity prevention for adolescents: A cluster randomized controlled trial in a school setting. *American Journal of Preventive Medicine*, *44*, 30–39.
- Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *British Medical Journal*, *299*, 313–315.
- Bugge, C., Williams, B., Hagen, S., Logan, J., Glazener, C., Pringle, S., & Sinclair, L. (2013). A process for decision-making after pilot and feasibility trials (ADePT): Development following a feasibility study of a complex intervention for pelvic organ prolapse. *Trials*, *14*, 353–366.
- Cambon, L., Minary, L., Ridde, V., & Alla, F. (2012). Transferability of interventions in health education: A review. *BMC Public Health*, *12*, 497–510.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Hopewell, NJ: Houghton Mifflin Company.

- Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, *321*, 694–696.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716.
- Colsher, P. L., & Wallace, R. B. (1989). Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology*, *44*, 45–52.
- Craig, P., & Peticrew, M. (2013). Developing and evaluating complex interventions: Reflections on the 2008 MRC guidance. *International Journal of Nursing Studies*, *50*, 585–587.
- Department of Health and Human Services. (2009). *Federal coordinating council for comparative effectiveness research*. Report to the President and Congress. Washington, DC: Author.
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials*. London: Arnold.
- Dunn, G. (2002). The challenge of patient choice and nonadherence to treatment in randomized controlled trials of counseling or psychotherapy. *Understanding Statistics*, *1*(1), 19–29.
- Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, *28*, 1167–1187.
- Fernald, D. H., Coombs, L., DeAlleaume, L., West, D., & Parnes, B. (2012). An assessment of the Hawthorne Effect in practice-based research. *Journal of the American Board of Family Medicine*, *25*, 83–86.
- Fuller, D., & Potvin, L. (2012). Context by treatment interactions as the primary object of study in cluster randomized controlled trials of population health interventions. *International Journal of Public Health*, *57*, 633–636.
- Green, R. J., & Herget, M. (1991). Outcomes of systemic/strategic team consultation: III. The importance of therapist warmth and active structuring. *Family Process*, *30*, 321–336.
- Guevel, M. R., Pommier, J., & Jourdan, D. (2013). A health promotion intervention in French primary schools based on teacher training and support. *Global Health Promotion*, *20*, 13–19.
- Guthrie, E. (2000). Psychotherapy for patients with complex disorders and chronic symptoms. The need for a new research paradigm. *The British Journal of Psychiatry*, *177*, 131–137.
- Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: how “out of control” can a randomised controlled trial be? *British Medical Journal*, *328*, 1561–1563.
- Hawe, P., Shiell, A., & Riley, T. (2008). In response to Spillane V., Byrne M. C., Byrne M., Leatham C. S., O’Malley M. & Cupples M. E. (2007). Monitoring treatment fidelity in a randomized trial of a complex intervention. *Journal of Advanced Nursing* *60*, 343–352. Important considerations for standardizing complex interventions. *Journal of Advanced Nursing*, *62*, 267.
- Heiman, G. W. (2002). *Research methods in psychology* (3rd ed.). Boston, MA: Houghton Mifflin Company.
- Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists’ responses to training. *Journal of Consulting and Clinical Psychology*, *61*, 441–447.
- Higgins, J. P., & Altman, D. G. (2008). Assessing risk of bias in included studies. In J. P. T. Higgins, D. G. Altman, & J. A. C. Sterne (Eds.), *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0 [updated March 2011]. Retrieved from www.cochrane-handbook.org.
- Horvath, A. (2005). The therapeutic relationship: Research and theory. *Psychotherapy Research*, *15*, 3–7.
- Howard, I., Turner, R., Olkin, R., & Mohr, D. C. (2006). Therapeutic alliance mediates the relationship between interpersonal problems and depression outcome in a cohort of multiple sclerosis patients. *Journal of Clinical Psychology*, *62*, 1197–1204.

- Huisman, S., Maes, S., De Gucht, V. J., Chatrou, M., & Haak, H. R. (2009). Self-regulation and weight reduction in patients with type 2 diabetes: A pilot intervention study. *Patient Education and Counseling*, 75, 84–90.
- Jack, S. M. (2006). Utility of qualitative research findings in evidence-based public health practice. *Public Health Nursing*, 23, 277–283.
- Johansson, H., & Eklund, M. (2003). Patients' opinion on what constitutes good psychiatric care. *Scandinavian Journal of Caring Sciences*, 17, 339–346.
- Kannus, P., Parkkari, J., Niemi, S., Pasanen, M., Palvanen, M., Jarvinen, M., & Vuori, I. (2000). Prevention of hip fracture in elderly people with use of a hip protector. *New England Journal of Medicine*, 343, 1506–1513.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, 1–27.
- Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research*, 16, 161–172.
- Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., ... Keller, M. B. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology*, 71, 997–1006.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 64, 532–539.
- Legrand, K., Bonsergent, E., Latache, C., Empereur, F., Collin, J. F., Lecomte, E., ... Briançon, S. (2012). Intervention dose estimation in health promotion programmes: A framework and a tool. Application to the diet and physical activity promotion PRALIMAP trial. *BMC Medical Research Methodology*, 12, 146–160.
- Marchal, B., Westhorp, G., Wong, G., Van, B. S., Greenhalgh, T., Kegels, G., & Pawson, R. (2013). Realist RCTs of complex interventions – An oxymoron. *Social Science and Medicine*, 94, 124–128.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438–450.
- Minary, L., Cambon, L., Martini, H., Wirth, N., Acouetey, D. S., Thouvenot, F., ... Alla, F. (2013). Efficacy of a smoking cessation program in a population of adolescent smokers in vocational schools: A public health evaluative controlled study. *BMC Public Health*, 13, 149–159.
- Minary, L., Martini, H., Wirth, N., Thouvenot, F., Acouetey, D. S., Martinet, Y., ... Alla, F. (2009). TABADO: “Evaluation of a smoking cessation program among adolescents in vocational training centers”: Study protocol. *BMC Public Health*, 9, 411–417.
- Mayo, E. (1933). *The human problems of an industrial civilisation*. New York, NY: Macmillan.
- Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In N. Fulop, P. Allen, A. Clarke, & N. Black (Eds.), *Studying the organization and delivery of health services: Research methods* (pp. 188–220). London: Routledge.
- Meyer, J. (2000). Qualitative research in health care. Using qualitative methods in health related action research. *British Medical Journal*, 320, 178–181.
- Muller, L., & Spitz, E. (2003). Multidimensional assessment of coping: Validation of the brief COPE among French population. *Encephale*, 29, 507–518.
- Musick, M., Wilson, J., & Bynum Jr., W. B. (2000). Race and formal volunteering: The differential effects of class and religion. *Social Forces*, 78, 1539–1571.

- Mynors-Wallis, L. M., Gath, D. H., Day, A., & Baker, F. (2000). Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *British Medical Journal*, *320*, 26–30.
- Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal*, *332*, 413–416.
- Oswald, P. A. (2003). Does the interpersonal reactivity index perspective-taking scale predict who will volunteer time to counsel adults entering college? *Perceptual and Motor Skills*, *97*, 1184–1186.
- Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C., Hoy, C., ... Sparks, L. (2005). Natural experiments: An underused tool for public health? *Public Health*, *119*, 751–757.
- Porta, M. (2008). *A dictionary of epidemiology* (5th ed.). New York, NY: Oxford University Press.
- Power, R., Langhaug, L. F., Nyamurera, T., Wilson, D., Bassett, M. T., & Cowan, F. M. (2004). Developing complex interventions for rigorous evaluation—A case study from rural Zimbabwe. *Health Education Research*, *19*, 570–575.
- Ridde, V., & Haddad, S. (2013). Pragmatism and realism for public health intervention evaluation. *Rev Epidemiol Sante Publique*, *61*, S95–S106.
- Ridde, V., & Mak, G. (2009). La place des données qualitatives dans les évaluations en santé publique [Realist approach to the test of actual program evaluation]. *Actualité et dossier en santé publique*, *69*, 32–34.
- Ridde, V., Robert, E., Guichard, A., Blaise, P., & Van Olmen, J. (2012). *L'approche réaliste à l'épreuve du réel de l'évaluation des programmes* [Realist approach to the test of actual program evaluation]. *Canadian Journal Programme Evaluation*, *26*, 37–59.
- Roland, M., & Torgerson, D. J. (1998). What are pragmatic trials? *British Medical Journal*, *316*, 285.
- Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, *56*, 119–127.
- Schulz, K. F., & Grimes, D. A. (2002). Blinding in randomised trials: Hiding who got what. *Lancet*, *359*, 696–700.
- Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, *20*, 637–648.
- Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *71*, 452–464.
- Stiles, W. B. (1999). Signs and voices in psychotherapy. *Psychotherapy Research*, *9*, 1–21.
- Strupp, H. H., & Hadley, S. W. (1979). Specific vs nonspecific factors in psychotherapy. A controlled study of outcome. *Archives of General Psychiatry*, *36*, 1125–1136.
- Tessier, S. (2008). *Effets bénéfiques de l'activité physique sur la qualité de vie liée à la santé et la corpulence: une approche épidémiologique et évaluative* [Beneficial effects of the physical activity on health-related quality of life and the corpulence: Epidemiological and evaluative approaches] (Thèse). Université de Lorraine, p. 308.
- Tessier, S., Vuillemin, A., & Briançon, S. (2006). Faut-il respecter la clause d'ambivalence dans les essais avec randomisation collective? [Should we respect the clause of ambivalence in trials with randomization group?]. *Revue d'Epidémiologie et de Santé Publique* *54* (Hors Série II), 89.
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., ... Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology*, *10*, 1–10.
- The Cochrane Collaboration. (2004). *Cochrane reviewers' handbook 4.2.2* (1st ed.). Chichester: John Wiley & Sons.

- Torgerson, D., & Torgerson, C. (2008). *Designing and running randomised trials in health, education and the social sciences*. Basingstoke: Palgrave Macmillan.
- Treweek, S., & Zwarenstein, M. (2009). Making trials matter: Pragmatic and explanatory trials and the problem of applicability. *Trials*, *10*, 37–46.
- Trompette, J., Kivits, J., Minary, L., Cambon, L., & Alla, F. (2013). *Stakeholders' perceptions of transferability criteria for health promotion interventions: A case study*. Unpublished work.
- Valbak, K. (2004). Suitability for psychoanalytic psychotherapy: A review. *Acta Psychiatrica Scandinavica*, *109*, 164–178.
- Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health*, *94*, 400–405.
- Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*, 914–923.
- Wang, S., Moss, J. R., & Hiller, J. E. (2006). Applicability and transferability of interventions in evidence-based public health. *Health Promotion International*, *21*, 76–83.
- Weisberg, H. I., Hayden, V. C., & Pontes, V. P. (2009). Selection criteria and generalizability within the counterfactual framework: Explaining the paradox of antidepressant-induced suicidality? *Clinical Trials*, *6*, 109–118.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist*, *59*, 595–613.
- WHO. (2001). *Evaluation in health promotion: Principles and perspectives* (Vol. 3). Copenhagen: Author.
- Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Juni, P., Altman, D. G., ... Sterne, J. A. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *British Medical Journal*, *336*, 601–605.
- Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine*, *300*, 1242–1245.