American Journal of Clinical and Experimental Medicine
2013; 1(1): 5-15
Published online June 10, 2013 (http://www.sciencepublishinggroup.com/j/ajcem)
doi: 10.11648/j.ajcem.20130101.12

Studying group behaviour: cluster randomized clinical trials

Charles J. Kowalski, Adam J. Mrdjenovich

Health and Behavioral Sciences IRB, Office of the Vice President for Research, University of Michigan, Ann Arbor, USA

Email address:
chuckk@umich.edu(C.J. Kowalski)

To cite this article:
Charles J. Kowalski, Adam J. Mrdjenovich. Studying Group Behaviour: Cluster Randomized Clinical Trials. *American Journal of Clinical and Experimental Medicine.* Vol. 1, No. 1, 2013, pp. 5-15. doi: 10.11648/j.ajcem.20130101.12

Abstract: Cluster randomized trials (CRTs) are experiments in which clusters of persons, rather than the persons themselves, are randomized to receive one of the interventions being studied. The use of CRTs has been increasing in response to the attention being paid to pragmatic as opposed to explanatory clinical trials, comparative effectiveness research, and community health promotional activities. We describe and illustrate the use of CRTs in these and other applications. Special attention is paid to ethical challenges in the design of such studies, and to tools for facilitating the implementation of interventions found to be efficacious in the trial into everyday clinical practice or effective community-wide programs. We argue that while CRTs have many useful and valid applications, there can be times when their use should be precluded due to ethical constraints. Special vigilance is required in research carried out in developing countries, where villages often seem to be a natural choice for clusters, but considerations of 'standard of care' may lead to control villages receiving no care or services. Full-fledged randomized controlled trials are not required to show that people who are doing poorly because of living in squalid conditions without proper sanitation and health care will, in the absence of change, continue to do so.

Keywords: Clinical Trials, Pragmatic Trials, Comparative Effectiveness Research, Health Promotional Activities, International Research, Ethics, Implementation Science, Ecological Validity

1. Introduction

Cluster randomized trials (CRTs) are experiments in which clusters of persons, rather than the persons themselves, are randomized to receive one of the interventions being studied. These clusters may consist of husband-wife pairs, families, classrooms, schools, clinics, health plans, neighborhoods, or even entire cities, or towns, or countries. Many have noted increasing use of the CRT, e.g.[3 – 6]. Among the reasons for this are the increasing use of pragmatic as opposed to explanatory clinical trials[7], and the attention being paid to ecological validity, comparative effectiveness research[8], and community health promotional activities. Puffer et al[9] may be consulted to get a feeling for the kinds of CRTs one might expect to find in the medical research literature. They studied the risks of bias in 36 CRTs that were published in the British Medical Journal, Lancet and New England Journal of Medicine during the years 1997-2002. The numbers of clusters in these studies ranged from 6 to 270; and the total numbers of participants from 130 to 182,200. The topics addressed included school interventions to reduce risk factors for obesity, general practitioner feedback to increase aspirin use, and maternal education for early treatment of pediatric malaria. Similar reviews were given for non-therapeutic interventions[10], and for primary prevention trials[11]. Not incidentally, while some methodological improvement over time, both within and between these studies, was in evidence, practitioners, all too often, are still not properly accounting for within-structure correlation during data analysis, and not taking between-cluster variation into account when determining sample size requirements. It has been noted that design and analysis issues may have special ethical consequences for CRTs[4]. Traditional (frequentist) design and analysis issues have been discussed by[5, 6, 12]. Bayesian approaches are also available[13-15]. An extension of the CONSORT statement to CRTs has been made[16]. It is not our intention to review all of this material. We focus on situations in which CRTs are apt to be advantageous, paying special attention to ethical issues and tools for assuring that interventions found to be efficacious can be *implemented*, i.e., that they will prove to be effective when applied. The importance of this last step has been recently recognized by the creation of a new journal, *Implementation Science*, that has defined implementation

research as "the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality of health services and care." This journal can be accessed at www.implementationscience.com. Woolf[17] has situated implementation science as the second step of the more general "bench to bedside" aim of *translational research*,[1] whose importance is evidenced by the existence of three journals : *Journal of Translational Medicine* (www.translational-medicine.com), *Translational Medicine* (www.omicsonline.org/translationalhome/php), and *Clinical and Translational Medicine* (www.springer.com/medicine/journal/40169). The ethics of translational research has received much attention, e.g., two target articles, each with discussion, in the *American Journal of Bioethics*[19, 20].

2. General Considerations for CRT Usage

Often the use of a CRT will be dictated by the fact that the interventions being studied are delivered to and affect *groups* of people; an oft-cited example is the use of different teaching methods in several classrooms. It is clear that in such cases, students within classes are not independent and so CRTs will be less efficient than a corresponding randomized controlled trial (RCT) in which individuals are the units of randomization. Thus it is incumbent upon designers of CRTs to explain the rationale behind the approach and, upon data analysis, to use methods that account for the within-cluster correlations among individuals. In addition to practical constraints like the teaching method/classroom example given above, reasons for employing CRTs are often commonsensical, having political and/or ethical components. For example, an intervention targeted at health care professionals involving the use of different guidelines for treating a given condition, would quite naturally involve a given physician using one set of guidelines on *all* her patients. Raudenbush thought that, in many situations, "results of demonstration projects based on random assignment of clusters may generalize better to the policy contexts they are designed to inform. The unit of assignment and treatment – the cluster – is often the unit of reform. After the research is completed, a preschool or a therapist will adopt a new approach for all clients, not for a randomly selected half"[12, p. 173]. Otherwise stated, when an intervention will be eventually applied at the cluster level, a CRT has the advantage of external validity. It is also true that individual benefit from a trial can be enhanced when others in the same cluster also benefit, and CRTs are uniquely positioned to characterize these sources of benefit. For example, it has been noted that, "… many trials in developing countries are of interventions against infectious diseases; here CRTs can measure the overall effect of an intervention on an individual's susceptibility to infection and the indirect effect due to changes in risk transmission to other individuals or herd immunity"[3, p. 4]. Another example where the use of clusters may prove advantageous is in avoiding the so-called *contamination problem*: If individuals are randomly assigned to treatments, but are members of existing clusters, others in the cluster may learn about and adopt the treatment that was *not* assigned to them, and estimates of treatment difference would tend to be understated (For example, when studying the effectiveness of a diet, members of the control group might learn of the diet and adopt it). Torgerson[21] thought that contamination would not be of much practical importance in many situations, but it is an issue that should not be ignored. As is generally true, "The starting point for experimental design should always involve joint consideration of the aims of the investigators and the constraints of the situation"[2, p. 164]; when practical constraints dictate that interventions be administered to groups rather than individuals and/or results will be applied en masse, the use of a CRT will often be appropriate. We might also note that the use of CRTs will more often be appropriate for pragmatic, rather than explanatory, clinical trials[7]. For example, all the patients of a given physician may form a cluster in order to better reflect real world, day-to-day clinical practice. This is the topic of the next section.

3. Pragmatic Clinical Trials

The differences between explanatory and pragmatic clinical trials can be summed up by: "explanatory trials measure *efficacy* – the benefit a treatment produces under ideal conditions, often using carefully selected subjects in a tightly controlled research setting; pragmatic trials measure *effectiveness* – the benefit a treatment produces in routine clinical practice, under real-world conditions, using representative samples of subjects"[7, p. 162]. Achieving 'real-world conditions in representative samples of subjects' is the motivation for comparing teaching methods in *classes*; campus improvement interventions in *schools*; productivity enhancers in *factories*; motivating healthy behaviors in *physicians' offices*; the effect of fluoride in the water supply in *communities*; the effects of community-based intervention studies in developing countries in *villages*. The idea is to achieve *ecological validity*: the setting of the study, and the methods and materials used, should correspond to the real-world setting, and the methods and materials to be used, when applied. It is seen then that CRTs may often be the method of choice in pragmatic clinical trials. They should be considered whenever the interventions are administered to groups and/or applied to them. There will be times, of course, when it will be more efficient to randomize individuals and

[1] Translational research is said to include two types of research: "T1, which moves basic science (bench) research to human clinical research (bedside), and T2 research which moves human clinical research (bedside) to clinical care (practice)"[18, p. 73]. We take T2 = Implementation science.

the traditional form of the RCT should be considered. It will, on occasion, be necessary to weigh the advantages and disadvantages of the two approaches. Multicenter clinical trials can be recognized as a form of CRT that might be so organized to mirror the real-world fact that the treatments being compared will, in practice, be administered in more than one place. In reality, of course, many multicenter trials are organized as such simply to satisfy sample size requirements: It may be infeasible to identify sufficient subjects from a single institution. The choice of design often is influenced by circumstance.

It should be noted that *observational studies*, where subjects are followed through the course of daily living in the real world is a very pragmatic approach to studying human behavior. These studies can differ in important ways from *trials* and special care in their design and analysis will be required[22]. However, in addition to ecological validity, observational studies can be very large and continued over extended periods of time, making them all but indispensible for the study of some topics, e.g., drug safety[23]. One setting in which drug safety is especially important is in comparing the risks and benefits of commonly used drugs. The comparative effectiveness of two drugs will depend not only on their respective potential benefits, but on their risk profiles, as well. One way to approach such a study is to use a CRT in which clusters are comprised of health plans. These organizations will have detailed background information on their members and (relatively) complete records of drug usage and side effects.

4. Comparative Effectiveness Research

Sabin et al[24] concentrated on the ethics and feasibility of CRTs in the context of comparative effectiveness research (CER) conducted by two or more health plans. They thought that while CRTs were "a cost-effective approach to comparing the effectiveness of commonly used drugs in representative groups of individuals," they were not widely used for this purpose because of (i) the traditional focus of RCTs on efficacy rather than effectiveness, (ii) the lack of financial support for CER trials, (iii) the small numbers of organizations with infrastructures able to support such trials, and (iv) concerns about the ethics of cluster randomization and the need for individual informed consent. We update these items and argue that the use of CRTs and health-plan-data offers fertile ground for CER.

With respect to the first item, while it is true that the traditional focus of RCTs is on efficacy rather than effectiveness[7], this focus is surely misplaced in the CER context. The Institute of Medicine (IOM) offered the definition: "CER is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition or to improve the delivery of health care … The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels." The classical randomized, placebo-controlled, double-blind clinical trial (RCT) has played a prominent role in testing the efficacy of new health care interventions, especially drugs, *but it is not designed to answer the questions posed by CER*. Luce et al[25, p. 206] noted that, "[a]s currently designed and conducted, many RCTs are ill suited to meet the evidentiary needs implicit in the IOM definition of CER: comparison of effective interventions among patients in typical patient care settings, with decisions tailored to individual patient needs." The typical placebo-controlled RCT does not directly compare effective interventions; it is not carried out in typical patient care settings; individual patient needs are subordinated to standardized treatment protocols; and randomization does not allow for patient choice or for treatments-tailored-to-individuals. Many RCTs employ simple outcome measures – or surrogates for these – ignoring quality of life outcomes which may be more important to patients and more relevant for policymakers[26, 27]. CRTs, however, need not be designed as "a typical RCT." Placebos are not an issue, as CER compares interventions already tested for efficacy. For another thing, the clusters are often comprised of groups of people, living and acting together as they do in the "real world."

Item (ii) has been addressed by the $1.1 billion that The American Recovery and Reinvestment Act of 2009 has allocated to support CER. Government officials have characterized the $1.1 billion as a "down payment" on a national program of CER[28, p. 203]. In addition, the Patient Protection and Affordable Care Act of 2010 established a private, nonprofit entity to oversee publicly financed CER, the Patient Centered Outcomes Research Institute (PCORI), whose core mission is to identify priorities for CER, fund these studies, and support improvements in CER methodology. For details, see www.pcori.org. Some think that public funding for CER may exceed $500 million per year by 2014[29].

Item (iii) was addressed by showing that many health plans were worthy candidates: "CRTs could be especially advantageous if implemented in the context of health plans that have extensive information on members, treatments, and clinical outcomes, together with an existing research infrastructure"[24, p. 40]. Sabin et al[24] focused on the situation in which A and B are widely used treatments (no placebos) for a common condition. Several health plans agree to participate in a CRT, some of the plans to "favor" A, the others, B ("favor" means that unless clinicians have specific reasons for choosing the non-preferred agent in a particular case, they will use the preferred one). This mirrors ordinary practice in that physicians are free to prescribe the alternate therapy if there are clinical reasons for doing so; and the patients are *real patients*, being treated in the *real world*. Sabin et al went so far as to say "there is no 'experiment'" here[24, p. 41]. In any case, the health plan data bases can be used to compare outcomes, taking into account whatever supplementary information (personal characteristics, side effects, compliance, etc) the plan routinely collects.

With regard to item (iv), cluster designs present ethical challenges for two primary reasons: (i) CRTs involve groups rather than individuals, and this raises questions about who may speak on behalf of a group under what authority, and (ii) the ethical implications of trials with *indirect* effects on subjects (i.e., trials for which the units of randomization, experimentation, and observation may differ) are not well understood[30]. In view of these challenges, questions have been raised about the degree to which existing ethical principles—of clinical research in general and RCTs specifically—may apply to CRTs, whether modification and careful application of such principles is sufficient from an ethical perspective, or whether CRTs call for entirely new ethical principles[31]. The literature seems to arrive at the conclusion that CRTs only partly fit within the current paradigm of research ethics, and that ethical principles devised for individually randomized trials apply to CRTs *with some specific adjustments*[3, 31]. The adjusted principles must address the issue of how cluster interests might be represented and assessed when individuals within clusters have competing interests (i.e., no individual member of a cluster should be disadvantaged by the cluster's participation in a CRT)[3].

Of the seven domains for ethical clinical research proposed by Emanuel, Wendler, and Grady[32], informed consent has received the widest attention in the literature on CRTs. Edwards et al[30] address the primary questions: Who consents to trial entry, how, and when? The answer depends on the level at which an intervention is delivered. Specifically, Edwards et al[30] distinguish cluster-cluster trials (in which the intervention is targeted at an entire group or community such that individuals cannot act independently) from individual-cluster trials (in which the intervention is targeted at individuals within clusters who can consent to treatment). Similarly, the MRC[3] distinguishes "Type A" interventions, which are received by an entire cluster from "Type B" interventions, for which individuals can consent without reference to other individuals. The key ethical issue is that, in the case of cluster-cluster trials or "Type A" interventions, it may be impracticable or impossible to obtain consent from all targeted individuals prior to random assignment, if at all. Considering that such designs do not allow for individual choice, the concept of individual consent "is not helpful"[3, p. 9]. This follows the fact that individuals within clusters who refuse participation in a cluster-cluster trial still may not be able to avoid exposure to the Type A intervention; thus, the purpose of individual consent is defeated[31]. Instead, the decision to enter a cluster-cluster trial is made by a guardian or "cluster representation mechanism"[3, p. 10]. The identity and responsibilities of guardians depend on the nature of the cluster and intervention. According to Hutton[4], a single cluster typically has several potential guardians, and a trial should not proceed unless consent is obtained from all guardians. The ethical principle here is that guardians must act in good faith; they should only volunteer their cluster when trial entry would be in the cluster's best interest based on considerations of distributive justice, utility, and equity[30]. In addition, guardians (a) serve as representatives or advocates for individuals within clusters, (b) remain active and informed as the trial proceeds, and (c) exercise caution to avoid conflicts of interest (e.g., guardians should be independent of the research team)[3]. Sabin et al[24, p.42] synthesize discussions of guardians and conclude that "some form of representative mechanism can be allowed to consent for entry of the cluster into a study, but the process requires careful safeguards and should be conducted in a transparent manner." To make the consent process more transparent, Donner and Klar[33] suggest that articles describing results of CRTs should provide information about (a) the identity of guardians and how they were selected, (b) opportunities for individuals to avoid the risks of an intervention delivered at the cluster level, and (c) the consent of individual participants in an individual-cluster trial. The Ottawa Ethics of CRTs Consensus Group[34] recommends that researchers conducting CRTs obtain consent from subjects in individual-cluster trials including professionals and service providers, prior to or as soon as possible after cluster randomization, before any data collection or intervention has taken place, unless a waiver is approved. Permission to enter the cluster on the part of a guardian with legitimate authority is not equivalent to individual consent and therefore should not negate this obligation.

Returning to the special case of CER using health plan clusters, it is important that health plans that participate, or that might participate, in CER make this fact known to members, including prospective members, so that there is at least some potential for choice in the matter. Sabin et al[24] argued that this was analogous to teaching hospitals declaring themselves to be teaching hospitals so that patients might exercise some choice in this regard.

5. Implementation Science

As mentioned earlier, implementation science aims to translate research results into every day clinical practice. To evaluate the feasibility of the translation process, one can use the RE-AIM (Reach, Efficacy, Adoption, Implementation and Maintenance) framework. RE-AIM was developed by Glasgow et al[35] to evaluate the potential for translation and public health impact of health behavior interventions. Brief descriptions of these five dimensions are as follows: *Reach* is an individual-level measure of participation referring to the percentage and characteristics of persons included in the study; *Efficacy* outcomes should include both positive and negative measures; and behavioral, quality of life, and participant satisfaction outcomes as well as physiologic endpoints; *Adoption* refers to the proportion and representativeness of settings that adopt the shown-to-be-efficacious program; *Implementation* refers to the extent to which the intervention is delivered as intended; and *Maintenance* refers to both individuals and settings adopting the program as routine and becoming part of the

everyday culture of the organization. The Implementation Science Team at the National Cancer Institute has elaborated on these definitions, providing, for each dimension, a list of items that need to be included in the assessment (see htpp://cancercontrol.cancer.gov.IS). The RE-AIM framework is not limited to use with CRTs, but since many health behavior interventions will be tested in *settings*, CRTs will often be the method of choice. In any case, RE-AIM can help plan and select samples, interventions, settings and agents in ways that make it more likely that the results can be implemented and/or replicated in further studies. The goal is the development of *disseminable* interventions. The RE-AIM framework can also be used to help select the design best suited to answer the study question posed. There are times when one or more pieces of the RE-AIM puzzle are already in place, e.g., the efficacy of an intervention in a particular place/time/setting/sample combination may have already been established. The question of whether the intervention will be effective in another venue may be more one of *implementation* or *adaptation* than one of starting-from-scratch efficacy, and may not require the trappings and rigid control of a RCT (CRT).

An example of the use of RE-AIM in a non-CRT setting was given by Shubert et al[36] who translated a research-based falls prevention intervention into a community-based program. An intervention of proven efficacy was delivered to individuals belonging to a single Senior Center with an eye towards implementation. Members at the Center were made aware of the availability of the program which was to be administered free-of-charge. The sorts of questions raised were:

- Reach – Would the target population attend?
- Effectiveness – What was the adherence and compliance to the program? Were there individual improvements in falls risk factors?
- Adoption – Would staff at the center adopt the program and offer it past the funding period?
- Implementation – What adaptations, including optimal frequency and duration, should be made to meet the community needs, still adhere to core elements and achieve similar outcomes?
- Maintenance – Would the program be sustained by the community partners?

These same questions would remain relevant should the study be expanded to compare different programs at a number of centers, i.e., to a CRT. More information would be expected to accrue than if a simple efficacy outcome was studied alone. One might, for example, decide that a somewhat less efficacious program would do better, overall, if it in fact reached more individuals. We next consider an example of a CRT done in this spirit. Meyer et al[37] used three of the RE-AIM factors (Implementation was not considered, and the study period was too short to assess Maintenance) in a three-arm CRT comparing smoking cessation interventions delivered in general practice settings. The three interventions consisted of brief advice by the practitioner (PRAC), individually tailored computer-generated letters (LETT), and the combination of advice and letter (COMB). *Reach* was measured by the number of interventions provided during the study period, *effectiveness* by smoking abstinence at a 12-month follow-up, and *adoption* by the proportion of practices participating in the study, Adoption of the three types of intervention were similar: PRAC (66.7%), LETT (76.9%) and COMB (66.2%). Reach was best for LETT, least effective for PRAC. This may have something to do with the fact that while physicians must give the advice personally, office staff may be put in charge of sending letters. This is consistent with the numbers seen for Adoption, where the decision to participate or not was made by physicians. The distributions of the measure of reach were highly skewed, with a large number of practices providing none or only a few interventions and a small proportion treating substantial numbers of patients. Further analysis showed that the major factor limiting reach of the interventions was the failure of practices to proactively offer the interventions rather than patients' willingness to participate. The computer-based intervention alone or in addition to conventional practitioner-delivered advice was more effective than advice alone. Looking at the results through a RE-AIM lens, the authors were able to note that, "The lower effectiveness of the tailored letters compared to the combination intervention was compensated by the slightly higher reach" and "By being less effective and associated with lower reach the brief advice intervention resulted in only one-third to one-half of the number of abstinent patients compared to each of the other conditions."[37, p. 131] Results such as these allow one to identify where improvement is needed. In this case, the apparent reluctance of some physicians to personally deliver the brief advice might suggest that a CPT, where P stands for preference, might be better than a CRT[38]. Allowing practices to choose their own approach can answer the question: Which intervention works best when chosen by those caring for the patients? One can imagine that a physician who wants to offer advice will do a better job of this than one who reluctantly agrees to do so.

As described above, implementation science (IS) can be used to speed the results of CRTs into practice. The relationship between IS and CRTs, however, can also be viewed the other way around – CRTs can be used to evaluate different methods of implementation. McKenzie et al[39] tested an intervention designed to increase adherence to clinical practice guidelines (CPGs) for treating acute lower back pain. They used a CRT involving 92 general practices, each practice with approximately 25 patients, for a total $N \approx 2300$. Those randomized to the intervention took part in face-to-face workshops that, among other things, pointed to the futility of taking simple x-rays[2] and to the importance of

[2] The CPG notes that diagnostic x-rays are rarely necessary: Plain x-rays for acute non-specific lower back pain are of limited diagnostic value, expose people to unnecessary ionizing radiation and provide no benefits in physical function, pain, or disability.

staying active. Those in the control had the usual access to the CPGs. They measured outcomes at both the practitioner and patient levels to see if the intervention affected physician behavior and if this results in benefits for patients, thus using a CRT to evaluate a new approach to implementation.

6. International Research

A particularly promising venue for the application of CRTs is in developing countries where rural landscapes are dotted with naturally occurring clusters, villages. For example, the MRC noted, "CRTs may be especially valuable in developing country settings, where in rural areas in particular the sense of community is strong, and community-level consent and cooperation are essential"[3, p. 4]. There is no denying the observation that CRTs appear to be a "good fit" in these settings, but we stress that just because a CRT *can* be done, does not always mean that that one *should* adopt this design: *The question being asked determines the appropriate research architecture, strategy, and tactics to be used – not tradition, authority, experts, paradigms or schools of thought*[1, p. 1636]. We cite two examples given by Lavery et al[40] where CRTs were in fact done in such settings, but a closer examination of *the relevant question to ask* in one, and *just what evidence must be brought to bear to disturb equipoise* in the other, raise some important questions about the approaches adopted. In a third example, we illustrate that when the underlying question fits, and when the RE-AIM framework is followed, a CRT can be ethically employed and produce rigorous results. We believe that a full and balanced discussion of CRTs requires not only showing examples of all the good things that can happen when they are done right, but also some appreciation for what can go wrong in situations where alternative designs may be more appropriate.

To set the stage for a discussion of these examples, we first recognize that research sponsored by developed countries but carried out in developing countries may raise special ethical concerns. For example, whose ethical standards should apply when cultural differences and disparities exist in access to health care between host and sponsor countries? Must subjects receive the best proven therapy even if it is not available locally? What obligations do researchers have after the completion of a study, and for what period of time? Several sets of principles/guidelines have been developed to help ensure the ethical conduct of research in such contexts. A number of recommendations offered in the Ottawa consensus statement[34] are particularly germane to cluster trials carried out in international settings. Specifically, it is recommended that investigators conducting CRTs should (a) justify the study intervention (benefits and harms must be consistent with competent practice), (b) justify the control condition (control subjects must not be deprived of effective care or programs), and (c) protect cluster interests by seeking cluster consultation (e.g., regarding ways in which protections might be enhanced for vulnerable subjects—whose involvement is needed to answer a study hypothesis but for whom autonomy and/or privacy may be compromised due to their position in the cluster—without impeding research that could benefit entire groups). According to Weijer et al[31], the application of beneficence to cluster trials, especially those conducted in developing countries, raises two primary ethical questions: (i) Are researchers obligated to provide subjects who are exposed to the burdens of data collection, but do not receive the intervention hypothesized to be the most effective, with more than usual care (and, if so, is it possible or practicable to meet this requirement when entire communities are assigned to control groups)?, and (ii) Is there an obligation to modify or stop a trial prematurely if an intervention appears unsafe or unexpectedly effective (and, in the latter case, is the expectation that the more effective strategy should be offered to all participants/communities)?[41]. The concept of clinical equipoise generally helps frame and address such questions in relation to RCTs. However, to the degree that clinical equipoise is thought to emerge from a fiduciary relationship between physician-researcher and patient-subject—neither of which might be involved in a CRT—the application of clinical equipoise to cluster trials may be somewhat unclear. Further, when it comes to international CRTs, ensuring that subjects are not exposed to any treatment which is known to be inferior to treatment available in clinical practice begs the question of, "Available where?"[31]. Emanuel et al[42] built on previous work where seven principles governing ethical clinical research were identified[32]. Here they applied the seven in the international context and added an eighth principle, *collaborative partnership*, with benchmarks:[3]

- Develop partnerships with researchers, makers of health policies, and the community
- Involve partners in sharing responsibilities for determining the importance of health problem, assessing the value of research, planning, conducting, and overseeing research, and integrating the research into the health-care system
- Respect the community's values, culture, traditions, and social practices
- Develop the capacity for researchers, makers of health policies, and the community to become full and equal partners in the research enterprise
- Ensure that recruited participants and communities receive benefits from the conduct and results of research
- Share fairly financial and other rewards of the research

The idea is that seeking the community's agreement and input helps avoid exploitation, and increases the chances that

[3] The (now eight) principles defined the sections in Lavery et al[40] into which their case studies were grouped. Both of the examples we consider are in the section on *Scientific Validity*, though other principles are also involved.

the research is important to the community, and may produce benefit there. These ideas were revisited and brought up-to-date by Emanuel et al[43]. Leaving no set of ethical principles behind, Emanuel and Weijer[44] argued that a forth principle should be added to the Belmont trinity,[4] namely *respect for communities*, and thought that this principle "is reasonably interpreted as conferring upon the researcher an obligation to respect the values and interests of the community in research and, whenever possible, to protect the community from harms"[p. 171]. Respect for communities is demonstrated when researchers fulfill their obligation to protect and empower social institutions, and abide by decisions of legitimate communal authorities[31].

The first example from Lavery et al[40, p. 105 ff] involved evaluating an intervention involving home-based treatment strategies for neonatal sepsis in India.[5] 86 villages with some 40,000 inhabitants were involved; 39 villages received the intervention, and 47 acted as controls. The researchers thought that since neonatal care was generally not available in these villages, a package of home-delivered interventions for neonatal care, including management of sepsis, was feasible to administer and could reduce the neonatal mortality rate by at least 25% in 3 years. This estimate was based, at least in part, on previously obtained results from home-based survival programs for managing pneumonia, diarrhea, and malaria in children which were instituted in the 39 intervention villages. Female health care workers were recruited and trained to manage neonatal sepsis by regularly visiting new mothers and urging them to take the child to a hospital if sepsis was suspected, or by administering antibiotics should hospitalization not be possible. The study was conducted in India, by Indian investigators, funded by the Indian government, and addressed a problem of great social importance to the Indian people. The protocol was developed with community participation and was approved by the Indian Academy of Pediatrics. There were, nonetheless, some residual ethical concerns. The two major concerns related to scientific validity and allowing the control villages to continue to face what was known to be the serious consequences of sepsis. Two reviewers provided their thoughts: Zulfiqar Bhutta[45] and Marcia Angell[46]. These are summarized in turn below. It will be seen that differences of opinion are possible even in these circumscribed areas; and that the question of *what question is being asked* is fundamental. At an even deeper level, one must consider what question *should* be asked.

Bhutta[45] did notice that "no attempt was made to improve the standard of care for the concurrent control population, which was already known to be poor"[p. 111], but thought that the 'gold standard status' of a randomized trial was sufficient justification for not considering possible alternative approaches which would have "significantly weakened the scientific credibility of the data." He thought that at the end of the trial, should the intervention prove effective, the control villages should be assured that it would be implemented there, but this amounts to an implicit acceptance of the study as designed.

Angell[46, p. 114] was less impressed with the design of the study. She noted that, even if one assumed the need for a CRT, the 'R' part of the CRT was missing: The villages selected for the intervention were those which had long-standing programs to treat certain infections in pregnant women and their newborns. This made these villages not comparable to the control villages which did not have these programs. This "invalidates the study from the outset." We think it also not an encouraging sign that Bhutta's concerns would be addressed – after all, long-standing effective programs for treating other infections were not implemented in these villages. Apparently the thought was to do the randomization *once*, but then repeatedly trying new programs in the 'intervention villages' with no thought given to exponential departure from comparability that would accrue at the continuing expense of the control villages.

But Angell's major concerns were even more fundamental. She questioned why the study even needed to be done: "There is no scientific question about whether antibiotics are effective for neonatal sepsis. We know that they are. So a trial is not required for that reason." She did, however, recognize that the researchers might have been interested in another question: "If the point is to see if village women can be trained to recognize sepsis and treat it, then the trial could be designed differently. Instead of leaving the 47 villages untreated, treatment on the control villages could be provided by more highly trained medical personnel. That would provide a benchmark against which to compare the performance of the village women." This transforms the trial into a non-inferiority trial[47] in which the aim is to establish that the village women will not perform worse than those more highly trained. Perhaps even better, the question might be phrased as one simply asking whether the village women can do it: "… there needn't be a control group at all. The performance of the village women could be monitored by the researchers in a small number of villages to see how well they do. Their training could be adjusted accordingly until they are shown competent to recognize sepsis and treat it." This example clearly shows the importance of asking the relevant question – the question will determine the appropriate design. As indicated earlier, this question should be identified in the context of a collaborative partnership with those that will bear the brunt of the research burden. It seems unlikely that the locals would choose to maintain the dysfunctional health system prevailing in the control villages if they were fully informed of the considerable (known) risks and arguable benefits of doing so. Angell[46, p. 115] summed up: "Not every health intervention requires a clinical trial. Sometimes we understand quite enough to

[4] Three principles were articulated in the Belmont Report: Respect for persons, beneficence, and justice.

[5] This certainly fits 'rural areas in developing countries' where the MRC[3] thought CRTs might be especially valuable: Of the billion people living in India, about 70% live in villages, and 62% depend on agriculture for their income[40, p. 105].

know that certain medical services are badly needed. What is required is not a clinical trial, but the political will and the resources to provide the care."

The second example[40, p. 116 ff] is similar in many respects to the first, but here the focus is not on the question asked, but on whether or not clinical equipoise was in place at the study's inception. The study was conducted by researchers from an American university, sponsored by an international tropical-disease research program, and carried out in two areas with high malaria rates in Tigray, Ethiopia.[6] 24 clusters of villages were organized into 12 pairs; and then the members of each pair were randomly allocated to receive either an intensive community education program or to maintain the village's standard of care. Mothers in the intervention arm were taught to recognize the signs and symptoms of malaria and to promptly provide anti-malaria medication to their sick children in the home. Control mothers were expected to follow the Tigray Community-Based Malaria Control Programme (TCBMCP), a program that relied on trained, volunteer community health workers to expeditiously point affected children to the treatment they need, but whose apparent ineffectiveness prompted the research to find a better approach. The new intervention, naturally enough, built on the experiences of the TCBMCP. Extensive interviews and qualitative research within the communities helped shape the content of the new intervention. These efforts suggested the TCBMCP suffered from three major limitations: (i) there were but few community health workers in the region, (ii) they almost all lived in the main villages, and (iii) they were almost all male. By training mothers in each village to recognize malaria symptoms in their children, to give the appropriate course of chloroquine, to recognize possible adverse reactions to the drug, and to refer a child to medical care if they did not improve in 48 hours, the researchers thought that the intervention was "feasible, affordable, and more effective than the existing methods used to prevent deaths from malaria in children under 5"[p. 117]. This was based on the recognition that mothers would be highly motivated and in the best position to perform these functions, and the qualitative research findings mentioned above.

Almost 14,000 children under the age of 5 were in the study; 7,294 in the control arm, and 6, 383 in the intervention arm. The primary outcome was mortality attributable to malaria. The control arm suffered almost twice as many deaths per 1,000 children under 5 years of age as did the intervention arm, and the ethical question centers around whether these deaths were preventable. The two reviewers of the study agreed that the qualitative research done prior to the CRT was not sufficient to disturb equipoise, and thus that the CRT was ethically justifiable.

Lavery[48] summarized his thoughts as follows: "…despite the uncontestable value of the qualitative research in shaping the intervention, the Tigray trial also illustrates a fair judgment on the part of the investigators that the resulting design of the intervention would not have been implemented by public-health practitioners or supported by policy makers within the governments of poor countries without the 'higher order' evidence from the randomized trial"[p. 125]. We believe what this more likely does in fact illustrate is a failure of the researchers to develop a *collaborative partnership*[42], whose first benchmark deals with the partnership developed between the researchers, *makers of health policies*, and the community. This would include such questions as whether or not policy makers would support and the community would embrace the intervention in the absence of randomization. Researchers may be motivated more by their desire to publish in a prestigious journal than the locals who might be expected to want to repair programs that are broken, especially when a continuance of the broken program is likely to result in infant mortality. Putting emphasis on the power of the study to convince also puts one in the awkward position of realizing that the poorer the control performance, the more control children die, the more convincing the result will be.

Singh[49] thought that "Because qualitative research is perceived, rightly or wrongly, as lacking the evidentiary weight of the randomized clinical trial, and given the extraneous factors that could potentially have affected the outcome of the study, I do not believe that the investigators reasonably foresaw or should have foreseen the outcome of the study before the trial began. Accordingly, as the investigators were in a state of equipoise at the study's commencement, the study was not unethical on that basis"[p. 129]. We find it interesting that this judgment is independent of whether the perceptions of qualitative research are right or wrong. One might think that if qualitative research results *should* be taken seriously, a commentary on the ethics of the study might point out that beneficence requires that subjects be protected from the harm certain to befall those who will continue to be subject to a broken program. The real question seems to be that of *how to fix a broken program*. A strong argument can be made in many cases that a good approach is to speak to those who took part in the program about why the program failed and how the revealed shortcomings might be overcome. The fact that this is called *qualitative research* in no way detracts from the idea that this is the best way to answer this question.

The injunction to protect the community from harms – *respect for communities* – warrants emphasis in the Tigray malaria study. It had already been recognized that the TCBMCP had serious limitations and it's not much of a stretch to think that, if continued, children would continue to die. Including a control arm in the study was apparently justified by the contention that the efficacy of the proposed program had not been proven (its potential advantages had been pointed to only by qualitative research methods, not an all-powerful RCT). This, however, is one of those situations in which the hierarchy of evidence that confers gold-standard-status on randomized controlled trials[22]

[6] Again the 'rural areas in developing countries' description applies: Agriculture employs some 85% of the population of Ethiopia.

might be challenged. The efficacy of the anti-malaria drugs is not an issue. The old program wasn't working and the program's deficiencies were recognized and corrected. The question wasn't so much of whether the reworked program would be better, but more of whether the new program could be implemented. A CRT where one of the arms receives an intervention *known to not work* is not required to answer the question of implementation. When the CRT was carried out, twice as many deaths due to malaria occurred in the control arm. While we agree with the notion that a study is ethical or not at its inception, this result is troubling because of the dire consequences in children which should have been entirely predictable. This was a clear and present risk at the start of the study and should have figured more prominently in the risk/benefit calculus. As Hutton[4] noted, "Thorough analysis of the risks and benefits[of CRTs] is no less necessary than for RCTs. There is a temptation to think that preventive or educational measures[such as the intervention arm of the Tigray trial] carry no risks"[p. 485].

Our final example is one in which a CRT is aligned with the purpose of the study and steps on fewer ethical toes. Bolton et al[50] conducted a CRT of group-based interpersonal psychotherapy (IPT) in rural Uganda with twin objectives:

- To test the efficacy of group IPT in alleviating depression and dysfunction, and
- To evaluate the feasibility of conducting controlled trials in Africa.

These involve an E and an I in the RE-AIM framework: the E refers to IPT; the I to both IPT and to the controlled trials themselves – can psychotherapy trials be implemented in Africa? For the trial, 30 villages were randomly selected and assigned to receive either IPT for 16 weeks or usual care, i.e., individuals from control villages were free to seek out whatever interventions they wished, for example, treatments by traditional healers. They clearly stated that their purpose was not to replace existing methods for treating depression, but rather to see whether or not IPT would be a useful addition to available approaches. Thus, "the trial comparison is not IPT vs. nothing, but IPT vs. the usual treatment, whatever that may be"[50, p. 3123]. The authors were careful to note that "While there is substantial evidence for the efficacy of 'talking therapies' these have been developed[and evaluated] in industrialized nations in the Western Hemisphere. The extent to which the concepts[7] and therapeutic strategies they use are appropriate among other populations is unknown"[p. 3118]. They argued that the focus on E could not be replaced by a simple look at A in the RE-AIM framework: "In sub-Saharan Africa, conditions are very different from those in which psychotherapy was developed, in ways that could reduce effectiveness. For example, many populations are reluctant to communicate directly about sensitive issues; others live in conditions of extreme chronic deprivation that are rare in developed countries." This is not to say the 'A' was ignored – the adaptation of IPT to a culturally acceptable form for use in Uganda was a considerable challenge (described elsewhere) – but this was recognized in this context to be a consideration separate from that of efficacy. In any event, using a culturally appropriate questionnaire, diagnosis of depression was confirmed prior to the intervention and reassessed post intervention, at which time 7% of subjects in the experimental arm met criteria for depression versus 55% in the control arm. The intervention was thus shown to be effective, but the authors recognized that they were uncertain about how long the effects of IPT might last. They acknowledged the need to extend the study to assess this, and allowed for the possibility that it might be necessary to add a maintenance (M) component to prevent recurrence of depression.

7. Conclusions

Aspects of the Bolton et al[50] study serve to illustrate ways in which the potential challenges and ethical issues involved with cluster trials may be managed effectively. Design and analysis issues (e.g., sample size calculation, allowance for within cluster correlation and between cluster variability, and adjustment for nonindependence of outcomes) were addressed as described in the report[50]. A justification for the study intervention (i.e., that depression is common and serious in sub-Saharan Africa and few treatments are available for depressed persons in impoverished countries) and a description of community leaders who identified potential participants is also provided. Consent was obtained from individual participants prior to cluster randomization, and again before any data collection or intervention took place. IPT was made available to control subjects at the conclusion of the study.[8] Respect for communities, collaborative partnership, and cluster consultation are reflected in the (a) use of native terms for depression (b) cross-cultural adaptation and validation of measures (i.e., a depression checklist was revised using ethnographic methods such that items were based on tasks reported as being important to local people), and (c) segregation of groups by sex. Thus, the Bolton et al study illustrates that there will be times when a CRT may be appropriately applied amid the many constraints imposed upon research conducted in such contexts. This will not always prove possible. Nor will it always be necessary. Other study designs may be suggested by the nature of the question being asked and the constraints of the situation.

[7] In another, but related context, Hunt[51, p. 63] noted, "The notion of quality of life is, itself, a cultural construct, introduced originally by social scientists in the USA. It cannot be assumed to have universal relevance or meaning."

[8] We did note a potential problem in terms of the recruitment/selection strategy. A list of eligible subjects was created for each participating village. After 8 subjects were enrolled from a given village, no contact was made with persons whose name remained on the list. Thus, it appears at least some eligible subjects did not have the opportunity to receive group psychotherapy as a participant of the study.

References

[1] Sackett, D.L., and J.E. Wennberg. 1997. Choosing the best research design for each question: It's time to stop squabbling over the "best" methods. British Medical Journal 315: 1636.

[2] Lambert, M.F., and J. Wood. 2000. Incorporating patient preferences into randomized trials. Journal of Clinical Epidemiology 53: 163-66.

[3] MRC. 2002. Cluster randomized trials: Methodological and ethical considerations. MRC clinical trial series. Available at www.mrc.ac.uk

[4] Hutton, J.L. 2001. Are distinctive ethical principles required for cluster randomized controlled trials? Statistics in Medicine 20: 473-88.

[5] Klar, N., and A. Donner. 2001. Current and future challenges in the design and analysis of cluster randomization trials. Statistics in Medicine 20: 3729-40.

[6] Donner, A. 1998. Some aspects of the design and analysis of cluster randomization trials. Applied Statistics 47: 95-113.

[7] Kowalski, C.J. 2010. Pragmatic problems with clinical equipoise. Perspectives in Biology and Medicine 53: 161-73.

[8] Kowalski, C.J. 2013. Comparative effectiveness research: Decision-bases evidence making. Perspectives in Biology and Medicine

[9] Puffer, S., D.J. Torgerson, and J. Watson. 2003. Evidence of risk of bias in cluster randomized trials: Review of recent trials published in three general medical journals. British Medical Journal 327: 785-9.

[10] Donner, A., K.S. Brown, and P. Brasher. 1990. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. International Journal of Epidemiology 19: 795-800.

[11] Simpson, J.M., N. Klar, and A. Donner. 1995. Accounting for cluster randomization: A review of primary prevention trials, 1990 through 1993. American Journal of Public Health 85: 1378-83.

[12] Raudenbush, S.W. 1997. Statistical analysis and optimal design for cluster randomized trials. Psychological Methods 2: 173-85.

[13] Turner, R.M., R.Z. Omar, and S.G. Thompson. 2001. Bayesian methods of analysis for cluster randomized trials with binary outcome data. Statistics in Medicine 20: 453-72.

[14] Spiegelhalter, D.J. 2001. Bayesian methods for cluster randomized trials with continuous responses. Statistics in Medicine 20: 435-52.

[15] Matsuyama, Y., J. Sakamoto, and Y. Ohashi. 1998. A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial. Statistics in Medicine 17: 1893-1908.

[16] Campbell, M.K., D.R. Elbourne, and D.G. Altman for the CONSORT Group. 2004. CONSORT statement: extension to cluster 14 management trials. British Medical Journal 328: 702-8.

[17] Woolf, S.H. 2008. The meaning of translational research and why it matters. Journal of the American Medical Association 299: 211-3.

[18] Szilagyi, P.G. 2009. Translational research and pediatrics. Academic Pediatrics 9: 71-80.

[19] Maienschein, J., M. Sunderland, R.A. Ankeny, and J.S. Robert. 2008. The ethos and ethics of translational research. American Journal of Bioethics 8: 43-51.

[20] Sofaer, N., and N. Eyal. 2010. The diverse ethics of translational research. American Journal of Bioethics 10: 19-30.

[21] Torgerson, D.J. 2001. Contamination in trials: is cluster randomization the answer? British Medical Journal 322: 355-7.

[22] Kowalski, C.J. 2013. When ethics precludes randomization: Put prospective, matched-pair, observational studies to work. Perspectives in Biology & Medicine In press.

[23] Kowalski, C.J. 2013. Pharmacovigilance: Why watchful waiting will work. Theoretical Medicine and Bioethics In press.

[24] Sabin, J.E., K. Mazor, V. Meterko, S.L. Goff, and R. Platt. 2008. Comparing drug effectiveness at health plans: The ethics of cluster randomized trials. Hastings Center Report 35: 39-48.

[25] Luce, B.R., J.M. Kramer, S.N. Goodman, et al. 2009. Rethinking randomized clinical trials for comparative effectiveness research: The need for transformational change. Annals of Internal Medicine 151: 206-9.

[26] Kowalski, C.J., S. Pennell, and A. Vinokur. 2008. Felicitometry: Measuring the 'quality' in quality of life. Bioethics 22: 307-13.

[27] Kowalski, C.J., J. Bernheim, N.A. Birk, and P. Theuns. 2012. Felicitometric hermeneutics: Interpreting quality of life measurements. Theoretical Medicine and Bioethics 33: 207-20.

[28] Sox, H.C., and S. Greenfield. 2009. Comparative effectiveness research: A report from the Institute of Medicine. Annals of Internal Medicine 151: 203-5.

[29] Nellesen, D., H.G. Birnbaum, and P.E. Greenberg. 2010. Perspectives on comparative effectiveness research: Views from diverse constituencies. Pharmacoeconomics 28: 789-98.

[30] Edwards, S.J., D.A. Braunholtz, R.J. Lilford, and A.J. Stevens. 1999. Ethical issues in the design and conduct of cluster randomized trials. British Medical Journal 318: 1407-9.

[31] Weijer, C., J.M. Grimshaw, M. Taljaard, A. Binik, R. Boruch, et al. 2011. Ethical issues posed by cluster randomized trials in health research. Trials 12: http://www.trialsjournal.com/content/12/1/100

[32] Emanuel, E.J., D. Wendler, and C. Grady. 2000. What makes clinical research ethical? Journal of the American Medical Association 283: 2701-11.

[33] Donner, A. and N. Klar. 2004. Pitfalls of and controversies in cluster randomized trials. American Journal of Public Health

94: 416-22.

[34] Weijer C., J.M. Grimshaw, M.P. Eccles, A.D. McRae, A. White, et al. 2012. The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials. PloS Med 9: e1001346. Doi:10.1371/journal.pmed.1001346

[35] Glasgow, R.E., T.M. Vogt, and S.M. Boles. 1999. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. American Journal of Public Health 89: 1322-7.

[36] Shubert, T.E., M. Altpeter, and J. Busby-Whitehead. 2011. Using the RE-AIM framework to translate a research-based falls prevention intervention into a community-based program: Lessons learned. Journal of Safety Research 42: 509-16.

[37] Meyer, C., S. Ulbricht, B. Gross, et al. 2012. Adoption, reach and effectiveness of computer-based, practitioner delivered and combined smoking interventions in general medical practices: A three-arm cluster randomized trial. Drug and Alcohol Dependence 121: 124-32.

[38] Kowalski, C.J., and A. Mrdjenovich. 2013. Patient preference clinical trials: Why and when they will sometimes be preferred. Perspectives in Biology and Medicine

[39] McKenzie, J.E., S.D. French, D.A. O'Connor, et al. 2008. IMPLEmenting a clinical practice guideline for acute lower back pain evidence-based management in general practice (IMPLEMENT): Cluster randomized controlled study protocol. Available at www.implementationscience.com/content/3/1/11.

[40] Lavery, J.V., C. Grady, E.R. Wahl, and E.J. Emanuel, eds. 2007. Ethical issues in international biomedical research: A casebook. New York: Oxford University Press.

[41] Glanz, K., B. Rimer, and C. Lerman. 1996. Ethical issues in the design and conduct of community based intervention studies. In: Ethics and epidemiology, S. S. Coughlin and T. L. Beauchamp, eds. New York: Oxford University Press.

[42] Emanuel, E.J., D. Wendler J. Killen, and C. Grady. 2004. What makes clinical research in developing countries ethical? The benchmarks of ethical research. Journal of Infectious Diseases 189: 930-7.

[43] Emanuel, E.J., D. Wendler, and C. Grady. 2008. An ethical framework for biomedical research. In: The Oxford textbook of clinical research ethics, E.J. Emanuel, C. Grady, R.A. Crouch et al, eds. New York: Oxford University Press, pp. 123-35.

[44] Emanuel, E.J., and C. Weijer. 2005. Protecting communities in research. In: Belmont revisited: Ethical principles for research with human subjects, J.E. Childress, E.M. Meslin, and H.T. Shapiro, eds. Washington: Georgetown University Press, pp. 165 – 83.

[45] Bhutta, Z.A. 2007. Did the SEARCH neonatal sepsis trial violate the Declaration of Helsinki? In: Ethical issues in international biomedical research, J.V. Lavery, C. Grady, E.R. Wahl, and E.J. Emanuel, eds. New York: Oxford University Press, pp. 109-14.

[46] Angell, M. 2007. The SEARCH neonatal sepsis study: Was it ethical? In: Ethical issues in international biomedical research, J.V. Lavery, C. Grady, E.R. Wahl, and E.J. Emanuel, eds. New York: Oxford University Press, pp. 114-5.

[47] Kowalski, C.J. 2013. Clinical trials of new drug products: What gets compared to whom? Perspectives in Biology and Medicine

[48] Lavery, J.V. 2007. The challenge of clinical equipoise in the Tigray malaria trial. In: Ethical issues in international biomedical research, J.V. Lavery, C. Grady, E.R. Wahl, and E.J. Emanuel, eds. New York: Oxford University Press, pp. 119-26.

[49] Singh, J. 2007. Could the investigators forsee the outcome of the Tigray trial? In: Ethical issues in international biomedical research, J.V. Lavery, C. Grady, E.R. Wahl, and E.J. Emanuel, eds. New York: Oxford University Press, pp. 126-30.

[50] Bolton, B., J. Bass, R. Neugebauer, et al. 2003. Group interpersonal psychotherapy for depression in rural Uganda. Journal of the American Medical Association 289: 3117-24.

[51] Hunt, S.M. 1998. Cross-cultural issues in the use of quality of life measures in randomized controlled trials. In: Quality of life assessment in clinical trials, M.J. Staquet, R.D. Hayes, and P.M. Fayers, eds. Oxford University Press, pp. 51-67.