# Cluster Randomized Stepped Wedge Blinded Controlled Trials (CRSWBCT) In Comparative Effectiveness Research (CER) – Part I: Toward a Revision of CONSORT

**Francesco Chiappelli\*[1], Andre Barkhordarian[1], Gary Demerjian[1,2], Quyen Bach[1] and Vandan Kasar[1]**

[1]Evidence-based Decisions Practice-Based Research Network, UCLA School of Dentistry, Westwood (Los Angeles), CA, USA
[2]Center for TMJ and Sleep Therapy, Burbank (Los Angeles), CA, USA

## Abstract

The current state of translational science and comparative effectiveness research (CER) faces inherent barriers to accuracy, which can compromise the integrity of the study. Design flaws such as information bias, inferential bias, selection bias, and reporting quality of trials require careful assessment in order to appropriately compare and contrast evidence in pursuit of the best evidence base (BEB). Here, we describe a novel research design, the Cluster Randomized Stepped Wedge Blinded Controlled Trials (CRSWBCT), which consists of a type of pragmatic trial that attempts to minimize these biases by implicating a system in which each unit of study acts as both the experimental group and control group. The CRSWBCT design has an inherent ability to enhance statistical stringency by minimizing risk of bias, and, in contrast to the traditional parallel, run-in, or cross -over trials, preserves ethical equipoise. In the CRSWBCT, all the clusters begin the study in the placebo-control, and roll-out into the experimental treatment group in a systematic, sequential fashion that retains the stringency of random double- blind protocol. CRSWBCT, while more complex in terms of power and statistical inference, has greater validity and is the preferred design with respect to ethical, logistical, and financial considerations over traditional simpler trials for the modern and contemporary pursuit of patient-centered outcomes research and individual patient data collection and analysis. In the context of CER, the adoption of CRSWBCT requires a revision of the CONSORT10 checklist (e.g., CONSORT10-R) to evaluate the evidence base obtained by the CRSWBCT design.

## Introduction

Translational medicine/dentistry, also referred to as translational science in medicine/dentistry, or translational science in healthcare [1], is a discipline within biomedical and public health research that aims to improve the health of individuals and the community by translating findings into diagnostic tools, interventions, procedures, policies and education. Its two principal facets correspond to translational research, analyzing biopsy materials obtained from a given patient to better inform patient-targeted treatment intervention, and translational effectiveness, utilizing the best evidence base (BEB) for a given patient-targeted treatment intervention in specific clinical settings [1,2].

The systematic process of garnering the available evidence, of assessing its level and quality, and of obtaining the consensus for BEB as it emerges from the entire body of available literature responding to a specific PICOTS (i.e., the bibliome), is the process of research synthesis, which generates the systematic review as its scientific dissemination product. CER utilizes research synthesis and systematic reviews to generate and to disseminate new comparative evidence of effectiveness of diagnostic tests, treatment interventions, clinical procedures, or other diverse modes of health-care service. The consensus that emerges from the systematic review is often in need of lay-language translation for dissemination to a variety of stakeholders, and for incorporation in evidence-based revisions of clinical practice guidelines and policies.

One key aspect to ensure the validity of CER outcomes lies in the validation of the instruments and tools recommended for use in the establishment of the quality of the research evidence. AHRQ and the Cochrane Group have independently developed a methodology for estimating the risk of bias inherent to any given research study. We have discussed risk of bias in translational medicine as a systematic error of methodology that pertains to measurement or sampling (e.g., selection bias), a systematic defect of design that leads to estimates of experimental and control groups, and of effect sizes that substantially deviate from true values (e.g., information bias), or a systematic distortion of the analytical process that results in a misrepresentation of the data with consequential errors of inference (e.g., inferential bias). We revised and quantified the AHRQ Risk of Bias instrument, because the risk of bias, in its totality, can seriously adulterate the internal and the external validity of a clinical study, and must therefore be identified and systematically evaluated [3].

A critical aspect of this process of relies on the verification that the research design of each individual study in a given systematic review meets certain fundamental criteria of research designs. To ensure the reporting quality of trials to be incorporated in a CER protocol, the Consolidated Standards of Randomized/Reporting Trials checklist (http://www.consort-statement.org), and its 2010 revision (CONSORT10) [4] is recommended. In brief, CONSORT10 consists of a 25- item checklist designed to provide guidance for reporting simple randomized controlled trials. More complex designs, including cluster randomized trails [5] and non-inferiority trials need a supplementary checklist. It is our contention presently that CONSORT must be re-evaluated once more, particularly in light of the recently evolved

stepped wedge designs.

The CRSWBCT is a type of pragmatic trial, that is a trial whose purpose is to inform decisions about practice, in which all the clusters begin the study in the placebo-control, and roll-out into the experimental treatment group in a systematic, sequential fashion that retains the stringency of random double-blind protocol. Consequently, one envisages that, in order to accommodate CRSWBCT, CONSORT10-R will need to include how to report a series of random trials conducted with a series of roll-outs/roll-ins from placebo to experimental arms, and how to assess blinding and cluster randomization in step wedged design trials, which removes the effects of a placebo.

## Methodological Rationale

Stepped wedge trials are a relatively recent evolution in the science of clinical research designs. They are superior to the traditional parallel, run-in or cross-over designs in that every unit of study will receive both placebo and the experimental intervention during the course of the study, thus ensuring the equipoise principle. Therefore, in terms of statistical stringency, stepped wedge trials are superior because every unit can act as its own control, which permits individual patient data assessments. Similarly, in terms of statistical efficiency, stepped wedge trials are complex, and can be under-powered. Here, we discuss the hypothetical incorporation of stepped wedge designs in CER, whose statistical model can be summarized as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij}\theta + \varepsilon_{ijk}$$

where $\alpha_i$ representing the cluster effect, $\beta_j$ representing the step effect, $\gamma_{ij}\theta$ representing the fixed treatment effect – by – treatment by cluster by step interaction, and $\varepsilon_{ijk}$ representing the residual random error

At time zero, none of the clusters receives the experimental mindful meditation intervention: they all are given controlled breathing/relaxation. At each subsequent time-points, one cluster switches over; the order and sequence is random and double-blind, thus yielding CRSWBCT. By the end of the study, all clusters have switched from placebo to the mindful meditation arm, and all the patients are receiving the experimental intervention. Since the order of roll-out/roll-in is random, the length of time of treatment with the experimental mindful meditation intervention is random.

## Discussion

Data analysis in CRSWBCT entails an individual-level generalized linear mixed model (LMM) analysis for repeated measures that requires proper weighting when cluster sizes vary. Ideally, CRSWBCT start with clusters of equal sample size, but invariably patient drop-out over time leads to missing data, which must be dealt with by data imputation or other statistical means, lest analyses with clusters of different sizes impair stringency of within- cluster analyses and call into question power. CRSWBCT power can be estimated based on the strength of treatment effect, the number of clusters, the number of steps, the sample size per cluster per step, and the variance components, and rendered as:

$$power = \phi\left(\sqrt{\frac{\theta_A^2}{Var(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}}\right)$$

Where $\Phi$ is the cumulative standard normal distribution function and $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the standard normal distribution function, and $Var(\hat{\theta})$ the appropriate element from the weighted least squares analysis.

Loss of power is more likely from loss of measurement times than loss of randomization of time points. Comparison of the data points in the placebo period vs. the intervention period for each patient within each practice of each cluster – to determine the overall effectiveness of the intervention.

When no pre-intervention measurements are obtained, the effects of the intervention cannot be separated from any underlying temporal changes due to the placebo (i.e., bias of cross-contamination) for each individual patient within each cluster. Despite this limitation, CRSWBCT remains advantageous, compared to a traditional cross-over RCT design, which suffers from a similar contamination bias. CRSWBCT further imposes distinct practical implementation challenges, which stress the importance of double-blinding as indicated above to prevent the "expectancy bias", as

- contamination between patients on the intervention and those the intervention to commence,

- knowledge of information on the part of those assessing outcomes with respect to participants' status in the protocol.

An important choice in CRSWBCT is the number of clusters randomized at each time step, which should be at least $n_c=3$. Randomizing multiple clusters at each time point reduces the overall number of measurement times, and affects power, as noted above. Power can be partly, but not completely, recovered by adding additional measurement periods onto the end of the trial. Therefore, an adequate number of time points for ensuring power of CRSWBCT is $n_{tp}=8$-10. But, while additional monitoring periods increase power, the full power that could have been obtained in a parallel design with no delay can seldom be recovered. This is particularly the case when the time-points delays are relatively short (e.g., days, weeks). In certain cases, depending largely on the residual variance, monthly time intervals may not be sufficiently long to recover the full intervention effect as would be noted in a single interval. On the other hand, because CRSWBCT analyses rely on within -cluster information, such as individual patient data analysis, the time delay effect on power is ultimately minimal.

## Conclusion

CRSWBCT offer a number of opportunities for data analysis, particularly for modeling the effect of time on the effectiveness of the intervention, by incorporating data collection at each point where a new cluster rolls-out of placebo and rolls-in to receive the intervention. This design will facilitate testing of overall efficacy and effectiveness of a number of interventions in translational medicine and more generally speaking in translational science. Part II of this writing discusses certain practical applications and implications of CRSWBCT for mindful meditation interventions to favor the outcome of personalized dental care in TMD patients with DA.

CRSWBCT has enhanced statistical stringency because of the utilization of many patients in several practices within multiple clusters, randomization within clusters, and stepped wedge repeated measures. It is the most appropriate design in most cases of individual patient data acquisition and analysis, particularly, when the intervention is expected, based on BEB, to proffer more good than harm. Parallel designs, in which certain participants receive only placebo, are in this instance unethical and violate, as noted above, the equipoise principle. Thus, CRSWBCT

- Resolves ethical dilemma of withholding the intervention when not in equipoise;

- Solves logistical and financial problems associated with

simultaneous implementation;

   •   Permits detection of trends by increasing statistical power within and between comparisons and

   •   Enables consideration of the outcomes across multiple settings and ethnicities.

In brief, CRSWBCT is a clinical trial with a complexity factor several orders of magnitude higher than the traditional parallel run-in or cross- sectional randomized blinded trials. CONSORT10 is insufficient to verify the structure of the design. Considering the strengths and advantages of stepped wedge designs and CRSWBCT in particular, expectations are that they will increasingly be incorporated in CER systematic reviews, which further emphasizes the urgency and timeliness of revising CONSORT (i.e., CONSORT10-R).

## Acknowledgements

## References

1. Chiappelli F (2014) Fundamentals of Evidence-based Health Care and Translational Science. Springer–Verlag, Heidelberg.

2. Chiappelli F (2015) Comparative Effectiveness Research: New Methods, Challenges and Health Implications. NovaScience Publisher, Inc., Hauppauge, NY.

3. Barkhordarian A, Pellionisz PA, Dousti M, Moura J, Chiappelli F et al. (2013) Assessment of Risk of Bias in Translational Science. Journal of Translational Medicine 11:184.

4. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med. 7:3

5. Chiappelli F (2012) Cluster Randomized Controlled Trials (CRCTs) in Evidence-Based Dentistry. Dental Hypotheses 3:1.