Randomized Controlled Trials
Shortcomings & Alternatives
iCBT Leiden 2017

Casper Albers University of Groningen

www.casperalbers.nl c.j.albers@rug.nl Twitter: CAAL

Source: Wikimedia Commons, user Arch

Overview

1. Short introduction
2. What are RCTs
3. What's wrong with them
4. Alternatives
5. Recommendations

Thanks to: Nitin Bhushan, Trudy Dehue and Maaike Nauta (all UG)

What are RCTs

Many definitions, e.g. (Sibbald, 1998): *Randomised controlled trials are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome and for assessing the cost effectiveness of a treatment. They have several important features:*

- *Random allocation to intervention groups*
- *Patients and trialists should remain unaware of which treatment was given until the study is completed [..]*
- *All intervention groups are treated identically [..]*
- *Patients are normally analysed within the group to which they were allocated, irrespective of whether they experienced the intended intervention [..] (intention to treat)*
- *The analysis is focused on estimating the size of the difference in predefined outcomes between intervention groups.*

What's wrong with RCTs

If it is possible to carry out a RCT: this hands down is the universally best approach.

RCTs suffer from practical and ethical disadvantages – which is why they cannot always be used.

What's wrong with RCTs 1/2

Many potential risks, e.g.:

- Is it ethical to assign healthy persons to a medical treatment; or to give affected ones only a placebo?
- RCTs are not always feasible, e.g. when it is impossible to blind conditions
- Absense of double blinding $\rightarrow$ effect sizes exaggerated (Schultz et al, 1995)
- RCTs are more expensive than other designs
- RCTs only applicable for linear causal effects (Carey & Stiles, 2016)
- 'Does this treatment work' only interesting for homogeneous groups.
- . . .

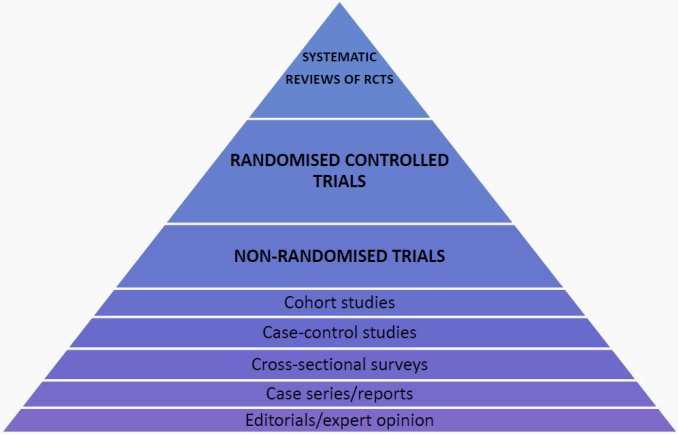
What's wrong with RCTs 2/2

- . . .
- Clear distinction IV/DV needed. In psychological processes, IV and DV influence each other ('principle of responsiveness')
- RCTs not suitable in psychological studies (too many variables contribute to psychological change; Carey & Stiles)
- Experiments can have low ecological validity. Experimental result $\neq$ evidence (Dehue, 2010)
- Many researchers don't deal with missing data properly (Zhang et al, 2017).
- RCTs only useful for evaluation purposes, not for discovery of new findings (e.g. Vandenbroucke, 2008).
- Also published RCTs suffer heavily from publication bias and other QRPs (Lancee et al, 2017).

Classification of studies

1. Fully Experimental Designs
   RCTs
2. Quasi-Experimental Designs
   Time series, uncontrolled before/after-studies, non-randomised controlled trials
3. Observational Designs
   Cohort studies, cross-sectional surveys

Classification of studies



Source: PPT by Dr. Mayston, King's College

Minimal requirements for causal claims

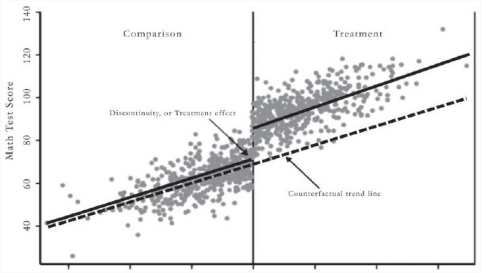
	Criteria	Definition
1	Strength	The size of the risk as measured by appropriate tests.
2	Consistency	The association is consistent when results are replicated in studies in different settings using different methods.
3	Specificity	When a single putative cause produces a specific effect.
4	Temporal sequence	Exposure always precedes the outcome.
5	Dose response	An increasing level of exposure (in amount and/or time) increases the risk.
6	Experimental evidence	The condition can be altered (prevented or ameliorated) by an appropriate experimental regimen
7	Biologic plausibility	The association agrees with currently accepted understanding of pathobiological processes.
8	Coherence	The association should be compatible with existing theory and knowledge.
9	Analogy	A finding of analogous associations between similar factors and similar diseases.

Source: PPT by Dr. Mayston, King's College

Alternative: regression discontinuity

Pretest-posttest quasi-experimental semi-causal design (cf. Panko et al., 2015).

1. Measure everyone's pretest score X_i (or some other score)
2. Assign those with $X_i < \theta$ to Group 0, with $X_i \geq \theta$ to Group 1
3. Model $Y_i = \beta_0 + \beta_1 \text{Group}_i + \beta_2(X_i - \theta) + \beta_3 \text{Group}_i(X_i - \theta) + \varepsilon_i$



Alternative: regression discontinuity

Advantage:

1. (Relatively) easy to do

Limitations:

1. Heavy dependency on statistical assumptions. You never know whether assumptions are valid.
2. Extrapolation to scores further from θ not possible.
3. Low statistical power.

I.e.: don't do this.

Alternative: propensity score matching (Rosenbaum & Rubin, 1983)

- In case random assignment is impossible.
- Correct for selection bias (as good as possible) by creating groups that are 'statistically similar' on a set of relevant variables.
- Multistep procedure:

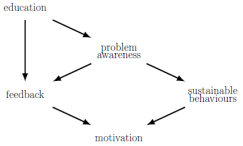
1. Run logistic regression, with $Y = 1$ in treatment and $Y = 0$ in control group
2. Result: propensity score
3. Check whether propensity scores are balanced across groups
4. Match each participant to a (or several) non-participant(s) with a similar propensity score
5. Verify that all controls are balanced
6. Run analysis on matched sample

Alternative: single case designs

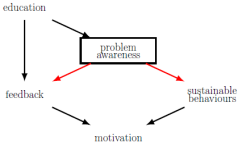
- RCTs study whether a treatment works *in general*.
- Often more interesting: does treatment work *for my patient*.
- $n = 1$: sample $\equiv$ population. No generalisation issues.
- Time series design advantage of temporal sequence.
- Studying states rather than traits
- Multiple simultaneous $n = 1$-studies: both within and between individual differences.
- Change in dynamics after starting treatment: *could be* treatment effect.

Alternative: Directed Acyclic Graphs

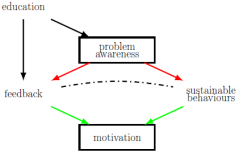
Especially useful for exploratory studies



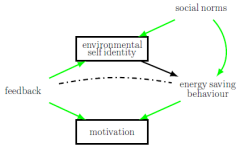
(a) A simple DAG, representing our theoretical knowledge underlying effects of environmental feedback on the household engaging in sustainable behaviours.



(b) Confounding bias illustrated using DAGs. Controlling problem awareness minimises spurious associations between feedback and sustainable behaviours (indicated by the red arrow).



(c) Collider bias illustrated using DAGs. Controlling motivation induces an association between feedback and sustainable behaviours (indicated by the green arrow).



(d) Causal mediation analyses, illustrated using DAGs

Source: Bhushan *et al*, 2017

Alternative: Directed Acyclic Graphs

Steps in constructing a DAG:

1. Design intervention based on theory
2. Draw a DAG with the processes underlying the intervention
3. Identify a sufficient set of factors which minimise bias
4. Conduct intervention, measure outcome variable and causal factors
5. Evaluate the intervention, controlling for causal factors

Controlling for a variable that is outcome of both the IV and the DV leads to collider bias

Alternative: Network Models

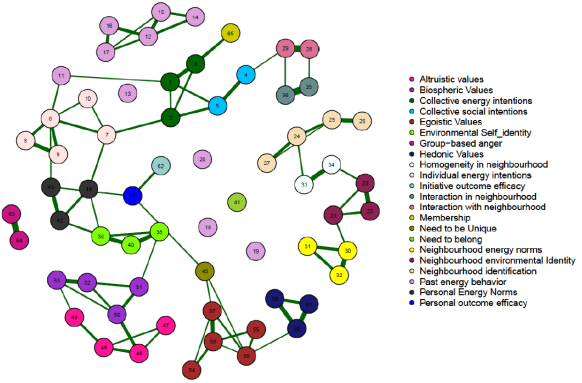


Figure 4: The GGM modelling conditional independence relationships between items.

Source: Bhushan, Mohnert *et al*, 2017

Some other alternatives

(From Carey & Stiles, 2016; and West *et al.*, 2017)

- Serial Replication
  Run a study multiple times under various settings (e.g. different countries, different time periods) and see what replicates
- Convergence of Evidence
  A single result is no result. Use a variety of sources.
- Benchmarking
  Don't compare your treatment with placebo (only); but with competing state-of-the-art treatments. Does it still outperform?
- Eliminating alternative explanations
- Randomized Encouragement Design
- Potential Outcomes Perspective
- Nonrandom Quantitative Assignment of Treatment

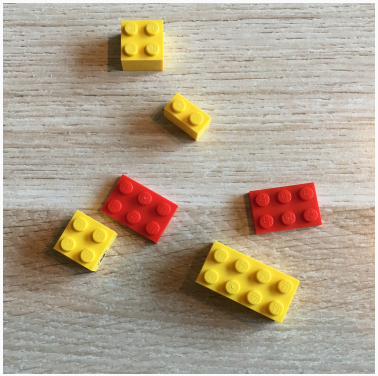
Recommendations by Hanin, 2017

1. Clinical trials should be publicly funded and conducted by [those with] no conflicts of interest.
2. Health care decisions based on outcomes of clinical trials should rely on a combination of statistical and biomedical evidence.
3. Scientific and health care benefits resulting from clinical trials should be compared to those of state-of-the-art controlled individual case studies incurring comparable costs.
4. Trials should be populated [with relevant participants].
5. Results of statistical analyses of randomized clinical trial data should be compared with those based on deterministic individual responses and permutation-based p-values, unless there is strong scientific evidence that individual responses are stochastic.
6. and 7. Justify your statistical choices (cf. Lakens *et al.*, 2017)

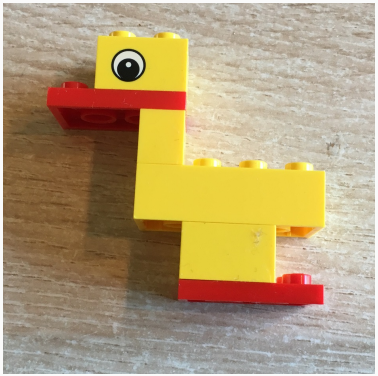
Recommendation by me

No single type of study is free from disadvantages.

Combine. Co-operate. Create.



Individual studies



Joint effort

References 1/2

- N. Bhushan, F. Mohnert, C.J. Albers, L. Jans, D. Sloot, E.M. Steg, 2017. Exploring the human dimension of the energy transition using graphical models. In preparation
- N. Bhushan, C.J. Albers, E.M. Steg, 2017. Behavioural Interventions and Directed Acyclic Graphs, submitted
- T. Carey & W. Stiles, 2016. Some problems with randomized controlled trials and some viable alternatives. Clinical psychology and psychotherapy.
- T. Dehue, 2010. Comparing Artificial Groups: on the history and assumptions of the randomised controlled trial. Ch.6 in: Will & Moreira (eds), Medical Proofs, Social Experiments, Routhledge
- M. Lancee *et al.*, 2017. Outcome reporting bias in randomized-controlled trials investigating antipsychotic drugs. Translational Psychiatry
- D. Lakens *et al.*, 2017. Justify your alpha: A Response to “Redefine Statistical Significance”, available from https://psyarxiv.com/9s3y6
- P.P. Panko *et al.*, 2015. Alternatives to RCT: A Review of Three Quasi-Experimental Designs for Causal Inference, Actualidades en Psicología, 29

References 2/2

- K.F. Schultz *et al.*, 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA
- P. Rosenbaum & D.B. Rubin, 1983. The central role of propensity score in observational studies for causal effects. Biometrika.
- B. Sibbald, 1998. Understanding controlled trials: why are randomized controlled trials important? BMJ
- J. Vandenbroucke, 2008. Observational Research, Randomised Trials and Two Views of Medical Science, PLoS Medicine
- S. West *et al.*, 2008. Alternatives to the Randomized Controlled Trial, Research Innovations and Recommendations
- Y. Zhang *et al.*, 2017. A systematic survey on reporting and methods for handling missing participant data for continuous outcomes in randomized controlled trials. Journal of Clinical Epidemiology

If you want more reading material on $n = 1$-models, network models or DAGs, please contact me