QUALITY OF STEPPED WEDGE TRIAL REPORTING CAN BE RELIABLY ASSESSED USING A NEW REPORTING GUIDELINE AND A CROWD-SOURCING REVIEW

NOVEMBER 2018

KARLA HEMMING INSTITUTE OF APPLIED HEALTH RESEARCH, UNIVERSITY OF BIRMINGHAM, UK.
BACKGROUND

- The Consolidated Standards Of Reporting Trials (CONSORT) extension for the Stepped-Wedge Cluster Randomised Trial (SW-CRT) is a recently published reporting guideline.
- The SW-CRT includes several design characteristics which make it different from the conventional cluster trial.
- Assessments of the quality of reporting of SW-CRTs against the CONSORT statement for CRTs have demonstrated poor reporting.
- Understandably, there have been no assessments of the quality of reporting according to the new CONSORT for SW-CRTs.

OBJECTIVES

▶ Assess the quality of reporting of a recent sample of SW-CRTs according to the newly developed reporting guideline.

  ▪ This assessment provides both a document of current reporting and will serve as a baseline assessment for any future study to identify any improvements over time.

▶ Determine if a crowd-sourcing type review is feasible to implement and reliable.

  ▪ If the crowd-sourcing methodology is reliable it has the potential to be used to our benefit at the current developments in cluster randomised trials and stepped wedge designs meeting.
OVERVIEW OF METHODS

- Identified the most recently published SW-CRT trial reports.
- Randomly allocated for quality assessment by participants who were attending *current developments in cluster randomised trials and stepped wedge designs meeting*, London, UK during November 2017.
- We expected approximately 50 participants to attend this workshop.
- To allow for independent extraction by two *(with content expertise)* reviewers per report, and to allow for the possibility of more participants than expected attending the workshop we sought to identify approximately 30 studies.
SCOPE OF REVIEW

- Included SW-CRTs:
  - Minimum of 3 sequences of allocations to periods spent in the control condition followed by periods in the intervention condition.
  - Two treatment conditions.
  - Cluster randomisation.

- Publication types:
  - Primary reports of SW-CRTs, i.e., protocols and reports of secondary analyses of a previously reported trial were excluded.
  - Open access or viewable from either the University of Birmingham or University of Ottawa libraries.
  - Published in English.
SEARCH PROCESS

- Objective was to identify the 30 most recently published SW-CRTs up to November 2017.

- We identified eligible studies in PubMed using a previously published search strategy [Martin 2016] run on the 21\textsuperscript{st} November 2017.

- We identified and ordered studies by date listed in Medline.

- To allow for exclusion of ineligible studies, titles and abstracts of the first 50 studies were screened in duplicate and independently by KH and MT and discussed to obtain a consensus on eligibility.

- Full copies of the reports were then obtained and assessed against the inclusion criteria, again in duplicate, identifying the required 30 full study reports.

CONSENT PROCESS

- All registered delegates invited by email to participate in the review.
- Participants were informed that attending the workshop would require undertaking a small amount of work in advance and during the event, with an invitation to contribute to the resulting manuscript as a *group author*.
- Anyone not wishing to participate was invited to opt out.
- Participants were also informed that data on inter-rater reliability as well as the quality of reporting would be evaluated (anonymously).
Participants not opting out were randomly allocated to one of the selected studies using computer-generated numbers, so that two participants were allocated to each study.

One week in advance of the meeting, a full PDF copy of their allocated study was e-mailed to participants, along with a simple quality assessment tool.

Participants were asked to independently assess their allocated study using this quality assessment tool.

Participants were kept blind to the other allocated assessor of the same report until the morning of the workshop.

After a 30 minute summary of the background to the project, the two participants assigned to each report met over a 30 minute period, discussed their discrepancies and reached a consensus.

Data were therefore abstracted independently and in duplicate.
Participants were asked to assess the quality of reporting for each of the 26 items according to a four-point scale:

- Clearly reported in full / clearly but partially reported / unclearly reported / not reported.
- Feedback after the independent assessment resulted in a change to a five-point scale with the addition of “not applicable” for the joint assessment.

The data were entered into an Excel database by one person (KH).
The inter-rater reliability:

- Percentage agreement (within item across pairs) and the Gwet A1 statistic using the `kappaetc` command in Stata 14.

- Using the four-point scale and by dichotomising the four-point scale into a two-point scale (clearly reported versus not clearly reported).

The joint assessment of the quality of reporting for all 26 items:

- Average number of items clearly or fully reported.

Excluded:

- Independent reviews if only one of the pair submitted their independent assessments (n=3).
- Any individual level items which were missing in either one or both independent assessments.
- Any assessments of not applicable because this was not included in the independent assessment tool.
Records identified through database searching: (n = 437)

Additional records identified through other sources: (n = 0)

Records after duplicates removed: (n = 437)

Records screened by date order: (n = 50)

Records excluded: (n = 22; incl. 7 not randomised; 7 protocols; 3 individually randomised; 2 not accessible; 1 methods and 2 secondary analysis)

Full-text articles assessed for eligibility: (n = 28)

Full-text articles excluded: (n = 1 secondary analyses and 1 individually randomised)

Studies included in the invitation to participate: (n = 26)

Full-text articles excluded by abstractors: (n = 1 individually randomised)

Studies included in the independent and joint quality assessment: (n = 25)
Participants registering for workshop:
(n = 53)

Participants invited to participate as reviewers:
(n = 50)

Participants invited to participate as reserves:
(n = 3)

Participants dropping out:
Before the day: n = 1
Duplicate registration: n = 1
Not attending on day: n = 2

Participants attending the workshop as reviewers:
(n = 47)

Waiting list invitation:
(n = 1)

Participants attending workshop as reserves:
(n = 3)

Independent assessments returned (un-blinded):
(n = 2)

Independent assessments returned (blinded):
(n = 45)

Joint assessments returned:
(n = 25)

Joint assessments not returned:
(n = 0)
RESULTS: RELIABILITY

Blue: Objective items? Red: Subjective items?
RESULTS: JOINT ASSESSMENT OF REPORTING

Who enrolled clusters / participants

Rationale for SW

Blinding

Harms

Protocol

Item number

0 20 40 60 80 100
CONCLUSIONS: REPORTING OF SW-CRTS

- To improve quality of reporting, authors of SW-CRTs should carefully ensure reporting of all the minimal items as described in the CONSORT extension for SW-CRTs.

- Particular attention should be given to ensure clear reporting on:
  
  • The exact format of the design with justification;

  • How the clusters and individuals were identified for inclusion in the study,

  • Whether this was done before or after randomisation of the clusters — both of which are crucial to the assessment of risks of bias.

CONCLUSIONS: THE FEASIBILITY OF USING CROWD-SOURCING

- Feasible and rated by last years participants as the best session at the meeting.
- Agreement between reviewers was low for some items.
- Possible explanations:
  - Wording in extension statement might be not so clear or irrelevant item (i.e. why the trial stopped)?
  - More training for consistency in reviewing or simply difficult to assess (i.e. “interpretation”)?
  - Or, inconsistency in reviewing common place just not so often assessed?

LIMITATIONS

- Convenience sample vs random sample of a well defined cohort (i.e. one calendar year)?
- Abstract screen was probably specific but lacked sensitivity:
  - For example, missed anything that was self defined as a wait list design.
- Small sample size (limited to 25 studies).
- Basic implementation.

More sophisticated systems
DISCUSSION

- Was the ethical oversight and consent process sufficient?
- What should we call the group authorship?
- Should we continue to use this approach in future meetings?
  - How to prioritise topics?
- Can the methodology be improved?
  - Assessment of reliability of joint assessment?
THE GOOD NEWS…

▶ Tentatively accepted by JCE (Journal of Clinical Epidemiology)
ACKNOWLEDGEMENTS

- K Carroll, J Thompson, A Forbes and M Taljaard and the SW-CRT review group
ACKNOWLEDGEMENTS
GROUP AUTHORS

- Susan Dutton OCTRU, University of Oxford susan.dutton@csm.ox.ac.uk; Vichithranie Madurasinghe QMUL v.madurasinghe@qmul.ac.uk; Katy Morgan LSHTM katy.morgan@lshtm.ac.uk; Beth Stuart University of Southampton bls1@soton.ac.uk; Katherine Fielding LSHTM katherine.fielding@lshtm.ac.uk; Victoria Cornelius Imperial College London v.cornelius@imperial.ac.uk; Liz Turner Duke University liz.turner@duke.edu; Richard Hooper QMUL r.l.hooper@qmul.ac.uk; Bruno Giraudieu INSERM U1246, France bruno.giraudieu@univ-tours.fr; Catey Bunce King’s College London ediz.hussein@kcl.ac.uk; Paul Seed King’s College London paul.seed@kcl.ac.uk; Alecia Nickless University of Oxford alecia.nickless@phc.ox.ac.uk; Michael Grayling MRC Biostatistics Unit mng21f@cam.ac.uk; Mélaine Prag Bordeau University, Inria, Inserm U1219, France melanie.prague@inria.fr; Sally Kerry QMUL s.m.kerry@qmul.ac.uk; Lauren Bell LSHTM Laurembell@lshtm.ac.uk; Eila Watson Oxford Brookes University ewatson@brookes.ac.uk; Rafael Gafoor King’s College London rahaf.gafoor@kcl.ac.uk; Nadine Martin QMUL n.martin@qmul.ac.uk; Emel Yorganci King’s College London Cicely Saunders Institute emel.yorganci@kcl.ac.uk; Lesley Smith Oxford Brookes University lesleysmith@brookes.ac.uk; Murielle Mbekwe INSERM U1246, Tours, France munellembekwe@yahoo.fr; Steven Teerenstra Radboud University Nijmegen z824116@umcn.nl; Claire Chan QMUL c.l.chan@qmul.ac.uk; Mirjam Moerbeek Utrecht University m.moerbeek@uu.nl; Pamela Jacobsen King’s College London Pamela.jacobsen@kcl.ac.uk; Simon Bond Cambridge Clinical trials Unit simon.bond@addenbrookes.nhs.uk; Agnès Caille INSERM U1246, France agnes.caille@univ-tours.fr; Ben Jones Plymouth University ben.jones@plymouth.ac.uk; Joan Preisser University of North Carolina, USA jpreiss@bios.unc.edu; Mona Kanaan University of York mona.kanaan@york.ac.uk; Catherine Hewitt University of York catherine.hewitt@york.ac.uk; Tracy Pellatt-Higgins University of Kent t.pellatt-higgins@kent.ac.uk; Laura Pankhurst NHS Blood and Transplant laura.pankhurst@nhsbt.nhs.uk; Schadrac Agbia LSHTM schadrac.agbia@lshtm.ac.uk; Sandra Eldridge QMUL s.eldridge@qmul.ac.uk; Robin Lerner University of Oxford robin.lerner@nhsoms.ox.ac.uk; Clémence Leyrat LSHTM clemence.leyrat@lshtm.ac.uk; Mark Pilling University of Cambridge mark.pilling@medscl.cam.ac.uk; Julia Forman King’s College London julia.forman@kcl.ac.uk; Indrani Bhattacharya ICR-CTSU indrani.bhattacharya@icr.ac.uk; Lesley Smith Oxford Brookes University lesleysmith@brookes.ac.uk; Nicholas Magill King’s College London nicholas.magill@kcl.ac.uk; Jane Candlish University of Sheffield j.candlish@gmail.com; Cliona McDowell NICTU cliona.mc dowell@nictu.hscni.net; James Martin Birmingham University J.T.Martin@bham.ac.uk; Louise Linsell National Perinatal Epidemiology Unit louise.linsell@npeu.ox.ac.uk; Caroline Kristunas University of Leicester cak21@le.ac.uk; Elizabeth Allen LSHTM elizabeth.allen@lshtm.ac.uk; Nadine Seward LSHTM nadine.seward@lshtm.ac.uk; Elaine Nicholls Keele University e.nicholls@keele.ac.uk; Bryony Franklin UCL School of Pharmacy bryony.deanfranklin@ucl.ac.uk;
THANK YOU