Published in final edited form as:

Lancet Glob Health. 2021 August ; 9(8): e1163–e1168. doi:10.1016/S2214-109X(21)00200-X.

Completeness of reporting and risks of overstating impact in cluster randomised trials: a systematic review

Elizabeth L. Turner, PhD [Associate Professor of Biostatistics and Global Health],
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA and Duke Global Health Institute, Duke University, Durham, NC, USA

Alyssa C. Platt, MA [Biostatistician III],
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA and Duke Global Health Institute, Duke University, Durham, NC, USA

John A. Gallis, ScM [Biostatistician III],
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA and Duke Global Health Institute, Duke University, Durham, NC, USA

Kaitlin Tetreault, MB [Research Student],
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA and Duke Global Health Institute, Duke University, Durham, NC, USA

Christina Easter, MSc [Medical Statistician],
Institute of Applied Health Research, University of Birmingham, Birmingham, UK

Joanne E. McKenzie, PhD [Associate Professor],
School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Stephen Nash, MSc [Research Fellow],

Corresponding author address: Elizabeth L Turner, Department of Biostatistics and Bioinformatics, Duke University, 11098 Hock Plaza, 2424 Erwin Road, Durham, NC 27705, USA. Tel: 919-681-6226. liz.turner@duke.edu.

Contributors

ELT and KH led the development of the project and wrote the first draft of the paper. AP developed the data abstraction tools with input from JG and ELT. All named authors (other than JM) conducted the initial abstract screen. ELT, ACP, JAG, and KT processed the data and conducted the statistical analysis. ELT and ACP invited participants and emailed relevant material. All group authors and named authors (other than ELT and KH) abstracted data from a published paper reporting a CRT included in this systematic review, participated in an in-person data reconciliation workshop, were invited to comment on the draft manuscript and all approved the final version. ELT is the guarantor.

Declaration of interest

All named authors have completed the *Unified Competing Interest form* (available on request from the corresponding author) and declare support from several grant funding organisations for the submitted work (see details in next section). The authors also confirm no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval

Ethical approval was not required as no human subjects research data was collected or used.

Trial registration details

The results reported in this manuscript are not based on a trial but instead based on summary statistics reported in multiple reports of cluster randomised trials.

Protocol

See Supplementary Material 1 (appendix pp 25-38).

PRISMA checklist

See Supplementary Material 4 (appendix pp 49-50).

Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

Andrew B. Forbes, PhD [Professor],
School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Karla Hemming, PhD [Professor]
Institute of Applied Health Research, University of Birmingham, Birmingham, UK

CRT Binary Outcome Reporting Group

Abstract

To avoid scale-up of interventions with smaller than perceived impact, complete and accurate reporting of expected impact is needed. This is of great importance in global health research to protect precious resources. In global health, the cluster randomised trial (CRT) design is commonly used to evaluate complex, multicomponent interventions and outcomes are often binary. Complete reporting of impact for binary outcomes means reporting both relative and absolute measures. We conducted a systematic review to assess reporting practices and potential to overstate impact in contemporary CRTs with primary binary outcome. Of 73 CRTs, most (82.2%) had incomplete reporting. Of 64 CRTs for which it was possible to evaluate, most (62.5%) were assessed as at risk of reporting in such a way that impact could be overstated. Care is needed to report complete evidence of impact for the many interventions evaluated using the CRT design around the globe.

Keywords

binary outcomes; cluster randomised trial; reporting; treatment effect measures; risk ratio; odds ratio; risk difference; rare outcome

Introduction

Well-conducted research is important in all settings, but particularly so in global health, where the accountability of researchers conducting "public health some where else” is ever more pressing.[1] Well-conducted randomised trials are required to evaluate the community-based, multicomponent and complex interventions that are so often of interest in global health research. Such interventions are typically evaluated using the cluster randomised trial (CRT) design.[2] Given the importance of these CRT evaluations for decision-making regarding the adoption and scale-up of interventions, it is important that impact be correctly estimated by accounting for clustering of outcomes and that impact be clearly communicated. For the binary outcomes commonly used in CRTs,[3] absolute (e.g. risk difference) and relative (e.g. risk ratio) measures are appropriate and provide complementary evidence. Their joint use has been recommended by the CONSORT statement on reporting of results of randomised trials, and subsequently of CRTs, for nearly two decades.[4-5]

The rationale for recommending the reporting of both absolute and relative effects has been long-documented.[6-8] First, there is the potential for large relative effects to mask those not of public health relevance and hence for the intervention effect to be overstated. This arises when outcomes are rare (e.g., < 10% risk), and no accompanying absolute measure is presented.[6-8] For example, in a CRT of an intervention to promote tobacco cessation in India, the reported relative risk of 5·32 (95% CI: 1·43 to 19·74) and absolute risk difference of 2·1 percentage points (pp) (95% CI: 0·7pp to 3·5pp) was obtained from a comparison of cessation proportions of 2·6% vs. 0·5%.[9] Such a magnitude of absolute difference could be deemed to be of insufficient public health benefit for the intervention to be scaled-up, yet had the relative effect been reported alone in this rare outcome setting, some users of the research findings may have overstated intervention impact. Second, there is the potential for the effect of an intervention to be overstated when an odds ratio is used as a relative measure of effect. This arises when outcomes are common (e.g., >10% risk) and the odds ratio is misinterpreted as a risk ratio.[10,11] For example, in a CRT of an intervention to link and retain adults in HIV care in Eswatini, the reported relative risk of 1·52 (95% CI: 1·19-1·96) was obtained from a comparison of proportions of 64% vs. 43%.[12] If the authors had reported an odds ratio in this common outcome setting, it would have been approximately 2·36. Had this been interpreted as a risk ratio, for example "the intervention led to a 136% increase in the proportion linked and retained in HIV care", it would be a considerable overstatement of impact given that the reported risk ratio estimate showed a relative increase of approximately 52%.

It is not yet known how many CRTs adhere to guidelines to report both relative and absolute measures, and reporting may be worse than for individually randomised controlled trials (RCTs) given the additional challenges of accounting for clustering in analysis of CRTs. A recent review showed that only 9% of individually-randomised RCTs presented both absolute and relative measures.[13] Given the importance of CRTs with binary outcome for evaluations in real-world contexts,[3,14,15] we undertook a systematic review of design, analysis and reporting practices of CRTs with primary binary outcome. Our goals were to determine the type of measures used, the completeness of reporting and the potential for impact to be overstated. Importantly, our investigation was not limited by country, context, condition or type of intervention but instead includes all identified CRTs to provide a full picture of global practice.

Systematic Review – Brief Methods

In brief, we included all reports registered in the Cochrane Central Register of Controlled Trials (CENTRAL) of two-arm parallel CRTs with at least one binary primary outcome that were published in 2017. See Table S1 (appendix p 8) for additional details and rationale regarding eligibility including the choice of 2017 and the restriction to parallel-arm designs and to full-scale CRTs (i.e. no pilot or feasibility CRTs). Data abstraction was undertaken in duplicate and by random assignment by more than eighty participants at three different CRT-focused workshops, all of whom were expert in statistical methods, and most expert in CRTs (Table S2, appendix 9). Final data was agreed by each pair and included consultation with lead study authors (ELT or KH), as needed. This strategy was adopted to enable the research to be conducted rapidly with high quality but without dedicated funding.[16] For more details, see the study protocol (appendix pp 25-38) and related materials (appendix pp 2-7, pp 39-50).

We report descriptive characteristics of the included studies (e.g., study domain, country, intervention type) and those undertaking the data abstraction. We summarize the proportion of studies reporting absolute and relative effects, as well as the proportion of studies that reported the primary outcome in such a way that there was potential for overstating the magnitude of the reported intervention impact. The latter was defined as either of the following two conditions: the outcome was rare (≤ 10% risk in either trial arm) and only a relative measure was reported; or, the outcome was common (> 10% risk in both arms) and the odds ratio was reported (see Introduction for examples of each of these situations). Our goal is to draw attention to how the choice of measures could lead to overstatement of impact by users of the research findings. Relatedly, in this article, we use the term "impact" to refer specifically to the reported effect estimates and recognize that the true "impact" is a function of implementation, compliance, heterogeneity of effects, sustainability over time, and transportability to other contexts.

Systematic Review - Key Findings

In brief, after removing duplicates or reports published only on a trials registration website, 711 abstracts were screened, from which 89 articles were eligible for full-text review. Data were abstracted by 82 individuals (Figures S1-S2, appendix pp 21-22; Tables S2-S3, appendix pp 9-10) from the 73 articles determined eligible to be included the review (Figure S3, appendix p 23; see a full reference list [appendix 51-54]). Comparisons of the pre-workshop independent data abstraction showed that inter-rater agreements between pairs exceeded 85% across the 82 individuals involved (Table S3, appendix p10) with a minimum of 83·6% at one workshop. Of the 73 CRTs in the review, the most common conditions studied were infectious diseases (19, 26·0%) and women's health (16, 21·9%), close to half (35, 47·9%) were conducted in at least one country designated as low or middle income (appendix p 19), with the remainder with all study sites in countries designated as high-income (33, 52·1%);[17] most randomised health facilities or providers (41, 56·2%) or by geographic area (14, 19·2%) and most (46, 64·8%) studied interventions targeted at participants, namely health promotion or educational interventions or direct participant therapeutic interventions (Table S4, appendix pp 11-12). Follow-up data were typically collected using a questionnaire or survey (34, 46·6%) or via electronic/medical records (22, 30·1%) (Table S5, appendix p 13). Most (52, 71·2%) enrolled fewer than 40 clusters and median (25th, 75th percentile) cluster size was 48 (20, 220) (Table S3, appendix p 10).

Few studies (13, 17·8%) reported both a relative and absolute measure and so, overall, most provided an incomplete picture of evidence of intervention impact (Figure 1, Table S6, appendix p 14). Instead, within the main text of each article, most (46, 63·0%) reported a relative measure only, 8 (11%) an absolute measure only and 6 (8·2%) reported no effect measure (typically reporting only proportions by arm, Figure 1) with a larger number (15, 20·5%) reporting no effect measure in the abstract. Of the 59 CRTs (80·8%) reporting a relative measure, most (40, 67·8%) reported an odds ratio, with fewer (19, 32·2%) reporting a risk ratio (Figure 1). Just 21 (28·8%) of the studies reported an absolute measure of effect (Figure 1), with most (19, 90.5%) reporting a risk difference and 3 (14.3%) reporting the number-needed-to-treat (1 of which also reported a risk difference).

Of the 64 CRTs reporting an effect measure with accompanying risks by arm, most (40, 62·5%) were classified as having the potential for users of the research to overstate intervention impact (Figure 1, Table S6, appendix p 14). Potential overstatement was primarily (28/40, 75%) because the odds ratio was the chosen relative measure for a common outcome, with the remaining 12 (25%) because only a relative measure (odds ratio or risk ratio or other) was reported for a rare outcome. The magnitude of this potential for overstatement is illustrated for the 59 studies that reported a relative measure (Figure 2). For the 28 CRTs in the common outcome setting that reported an odds ratio as the relative measure (shown in orange with reference risk > 10%), the odds ratio averaged about 40% larger than the risk ratio and, in one case, reached three-times the magnitude (Figure 2, footnote 3). Similarly, in the rare outcome setting with only a relative effect reported (shown in orange with reference risk ≤ 10%), those effects are typically of a large magnitude. For example, see the CRT with a risk ratio of almost 25 and a reference arm risk less than 5%.

Discussion

In all research, there is a duty to ensure that reporting of results of the expected impact of interventions is accurate so as to prevent poor use of precious resources. Our contemporary review suggests few cluster randomised evaluations with primary binary outcomes report both an absolute and relative measure of effect. As a consequence, most did not present evidence in a manner that facilitated accurate interpretation.[5] Incomplete, and potentially distorted, presentation of the evidence has implications for policy and practice; possibly leading to the adoption of interventions with smaller than perceived impact. Given the widespread use of CRTs in the evaluation of interventions in a range of clinical and public health settings across the globe, such limitations could have far-reaching consequences.

The finding that relative effects are typically reported alone may be a consequence of the inherent challenges in analysing the clustered binary outcome data that often arises in CRTs.[11,18] Indeed, we discovered that 9 (12.3%) of the 73 CRTs did not account for clustering in analysis (Table S7, appendix p 15). Yet, methods are available to estimate both risk ratios and risk differences for clustered data. These include binomial mixed-effects models or generalized estimating equations, with log (for risk ratios) or identity links (for risk differences), or cluster-level methods.[11] Relatedly, it may be surprising that relative effects were typically reported alone given that most (37, 78%) of the 48 journals in which the articles appeared endorsed the CONSORT statement on trial reporting (Table S8, appendix pp 16-17) and the majority (58, 79.5%) of the 73 CRTs appeared in one of those journals (Table S9, appendix p 18). As such, responsibility lies with journal editors to promote a shift in reporting practices. Likewise, there is a responsibility to avoid other poor practices identified in our review including: use of p-values in baseline tables; ignoring clustering in sample size calculations; and lack of clarity about the primary outcome or the primary assessment timepoint (Table S5, appendix p 13).

The finding that the odds ratio (OR) is the most commonly reported relative effect is likely explained by the fact that logistic regression mixed-effects modeling is the most well-known approach to analyse clustered binary outcome data.[18] Although ORs can be difficult to interpret and can lead to overstating of impact, they have several advantageous properties. First, whilst the absolute impact of an intervention may differ by underlying risk, ORs are often stable across underlying risk and therefore may be more applicable in different populations or different contexts.[19] Moreover, for outcomes that are very common (e.g. risk > 80%), the OR may be preferred to the risk ratio (RR) as a relative measure because the RR may mask the magnitude of some effects because absolute risk is bounded above by 100%.[20] Compared to their use in regular individually-randomised trials, there are additional challenges to interpretation of evidence from ORs in CRTs because intervention effects estimated by mixed models and generalized estimating equations using a logit link must be interpreted differently. That is to say, ORs estimated using mixed models should be interpreted as cluster-specific effects, whilst ORs estimated using generalized estimating equations should be interpreted as population-averaged effects.[2,15] The greater the variability between clusters, the greater the deviation of the cluster-specific odds ratio from the population averaged odds ratio. In contrast, when identity or log links are used to estimate risk differences and risk ratios, respectively, the population-averaged and cluster-specific treatment effects are identical.[11]

Potential limitations of our review include methodology and classification of overstatement. Regarding methodology, although we do not know how the relatively new “crowd-sourcing” approach would compare to a standard systematic review that relies on a few data abstractors, the “crowd-sourcing” approach has been previously used by our team to evaluate quality of reporting of stepped-wedge designs.[16] To promote high quality, data abstraction (appendix pp 39-48) focused on mostly objective measures, data abstraction was in duplicate by individuals with statistical expertise, of which most (65.9%, Table S2, appendix p 9) had CRT experience, and all pairs of data abstractors were able to consult with one of two senior authors during in-person workshops for data finalisation. Nevertheless, because we leveraged expertise from participants at UK and US-based workshops, few authors would be considered expert in global health research conducted in resource-constrained settings. Regarding our classification of overstatement, we focused on an objective measure of potential overstatement rather than actual misinterpretation, which would be very difficult to measure. As such, our estimate of 62.5% of CRTs with potential for overstatement of impact may be larger than what actually occurs in practice. In other considerations, whilst we limited ourselves to parallel-arm CRT designs, our findings likely extend to more complex designs including stepped-wedge, multiarm and crossover CRTs, as well as to observational studies with binary outcomes. Similarly, although we have not examined the nature of reporting of effects for safety outcomes or of findings from sub-group analyses, similar attention should be paid to avoid overstating impact in these important domains.

With the increasing move to use evidence generated from CRTs to make decisions regarding the adoption of interventions in settings around the globe, it is important that intervention impact be correctly and clearly communicated. For the binary outcomes so commonly used in decision-making, reporting of both relative and absolute measures of effect is necessary to provide complementary and complete information. Many researchers fail to realize the importance of this when communicating their study findings. Statisticians face some difficulties in estimating these effect measures in cluster randomised trials due to the complex nature of the models involved, but these are usually surmountable. Journals have a duty to ensure published trials adhere to consensus-based reporting guidelines and to ensure that both relative and absolute measures are reported. If these issues are not rectified, there will undoubtedly be negative consequences on the evidence used for decision-making about the adoption and scale-up of interventions around the globe.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Elizabeth Turner is funded in part by awards R01-AI141444 and R01-MH120649 from the National Institute of Allergy and Infectious Diseases and the National Institute of Mental Health, respectively, both part of the National Institutes of Health (US). John Gallis and Alyssa Platt's support of this project was made possible (in part) by Grant Number UL1TR002553 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Joanne McKenzie is supported by an Australian National Health and Medical Research Council Career Development Fellowship (APP1143429). Stephen Nash was supported by an award which is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, and part of the EDCTP2 programme supported by the European Union (grant ref: MR/R010161/1). Andrew Forbes acknowledges funding support from National Health and Medical Research Council of Australia Grant ID 1183303. Karla Hemming is funded by a NIHR Senior Research Fellowship SRF-2017-10-002. The contents of the research included in this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of any of the funders. The research contributed by all authors of this manuscript are independent of their funders. Specifically, the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors also wish to thank the Editor and three anonymous reviewers for their insightful comments and constructive feedback.

Funding

Partial funding from funding agencies in Australia (NHMRC), UK (DfiD, MRC, NIHR) and USA (NIH).

Data sharing

Data can be requested from the first author. Study protocol and statistical analysis plan are provided in supplementary material together with a dictionary outlining the database and corresponding fields.

CRT Binary Outcome Reporting Group

The following are members of the *CRT Binary Outcome Reporting Group* (see affiliations in appendix pp 55-56):

London, UK Workshop (with work location of each participant): Adrion, Christine (Germany); Akooji, Naseerah (UK); Bhangu, Aneel (UK); Bridgwood, Bernadeta (UK); Budgell, Eric (UK); Caille, Agnès (France); Campbell, Michael (UK); Chan, Claire Louise (UK); Collinson, Michelle (UK); Copas, Andrew (UK); Dixon, Stephanie N (Canada); Eldridge, Sandra (UK); Forster, Alice S (UK); Giraudeau, Bruno (France); Girling, Alan (UK); Glasbey, James (UK); Goulao, Beatriz (United Kingdom); Grantham, Kelsey L. (Australia); Hackett, Simon (UK); Hamborg, Thomas (UK); Hooper, Richard (UK); James, Kirsty (UK); Jarvis, Christopher (UK); Jones, Ben (UK); Kahan, Brennan (UK); Kanaan, Mona (UK); Kasza, Jessica (Australia); Kendall, Lindsay (UK); Kristunas, Caroline (UK); Leyrat, Clémence (UK); Macneill, S J (UK); Madurasinghe, Vichithranie (UK); Martin, James (UK); Mbekwe Yepnang, Ariane Murielle (France); Moerbeek, Mirjam

(The Netherlands); Mwandigha, Lazaro Mwakesi (UK); Ndounga Diakou (UK), Lee Aymar (France); Nepogodiev, Dmitri (UK); Omar, Omar (UK); Pankhurst, Laura A (UK); Perry, Hayley (UK); Rombach, Ines (UK); Stuart, Beth (UK); Taljaard, Monica (Canada); Tavernier, Elsa (France); Thompson, Jennifer A (UK); Wagner, Adam P (UK); Wilson, Nina (UK).

Durham, NC USA Workshop (all participants US-based) Cao, Shiwei; Harding, Monica; Kusibab, Kristie; Lee, Hui-Jie; McCormack, Kara; Moran, Kelly; Parish, Alice; Simmons, Ryan; Truong, Tracy; Vissoci, Joao Ricardo; Wang, Tongrong; Wang, Xueqi; Weber, Jeremy; Wilson, Jonathan; Yang, Siyun; Yang, Zidanyue.

Birmingham, UK Workshop (all participants UK-based) Bensoussane, Hannah; Bishop, Jon; Cheed, Versha; Gill, Alicia; Handley, Kelly; Hardy, Pollyanna; Hewitt, Catherine A; Ives, Natalie; Mehta, Samir; Patel, Smitaa; Sun, Yongzhong; Woolley, Rebecca

References

1. King NB, Koski A. Defining global health as public health somewhere else. BMJ Global Health. 2020;5:e002172.
2. Eldridge S, Kerry S. A practical guide to cluster randomised trials in health services research: John Wiley & Sons; 2012.
3. Fiero MH, Huang S, Oren E, Bell ML. Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. Trials. 2016;17(1):1. [PubMed: 26725476]
4. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332. [PubMed: 20332509]
5. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661. [PubMed: 22951546]
6. Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. J Gen Intern Med. 1993;8(10):543–8. [PubMed: 8271086]
7. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ. 2004;171(4):353–8. [PubMed: 15313996]
8. Perneger TV, Agoritsas T. Doctors and patients' susceptibility to framing bias: a randomized trial. J Gen Intern Med. 2011;26(12):1411–7. [PubMed: 21792695]
9. Sarkar BK, West R, Arora M, Ahluwalia JS, Reddy KS, Shahab L. Effectiveness of a brief community outreach tobacco cessation intervention in India: a cluster-randomised controlled trial (the BABEX Trial). Thorax. 2017;72(2):167–73. [PubMed: 27708113]
10. Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. N Engl J Med. 1999;341(4):279. [PubMed: 10413743]
11. Gallis JA, Turner EL. Relative measures of association for binary outcomes: Challenges and recommendations for the global health researcher. Annals of global health. 2019;85(1):137. [PubMed: 31807416] See also: Corrigendum: Relative measures of association for binary outcomes: Challenges and recommendations for the global health researcher. Annals of global health. 2020;86(1):77. [PubMed: 32704482]
12. McNairy ML, Lamb MR, Gachuhi AB, Nuwagaba-Biribonwoha H, Burke S, Mazibuko S, et al. Effectiveness of a combination strategy for linkage and retention in adult HIV care in Swaziland: The Link4Health cluster randomized trial. PLoS Med. 2017;14(11):e1002420. [PubMed: 29112963]
13. Rombach I, Knight R, Peckham N, Stokes JR, Cook JA. Current practice in analysing and reporting binary outcome data—a review of randomised controlled trial reports. BMC Med. 2020;18:1–8. [PubMed: 31898501]

14. Ivers N, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. BMJ. 2011;343:d5886. [PubMed: 21948873]

15. Hayes RJ, Moulton LH. Cluster randomised trials: Chapman and Hall/CRC; 2017.

16. Hemming K, Carroll K, Thompson J, Forbes A, Taljaard M, Dutton SJ, et al. Quality of stepped-wedge trial reporting can be reliably assessed using an updated CONSORT: crowd-sourcing systematic review. J Clin Epidemiol. 2019;107:77–88. [PubMed: 30500405]

17. Organisation for Economic Co-operation and Development. Classification of low and middle income countries. Available at: https://wellcome.org/grantfunding/guidance/low-and-middle-incomecountries. Last accessed March 18, 2021.

18. Li B, Lingsma HF, Steyerberg EW, Lesaffre E. Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. BMC Med Res Methodol. 2011;11(1):77. [PubMed: 21605357]

19. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med. 2002;21(11):1575–600. [PubMed: 12111921]

20. Cook TD. Advanced statistics: up with odds ratios! A case for odds ratios when outcomes are common. Acad Emerg Med. 2002;9(12):1430–4. [PubMed: 12460851]

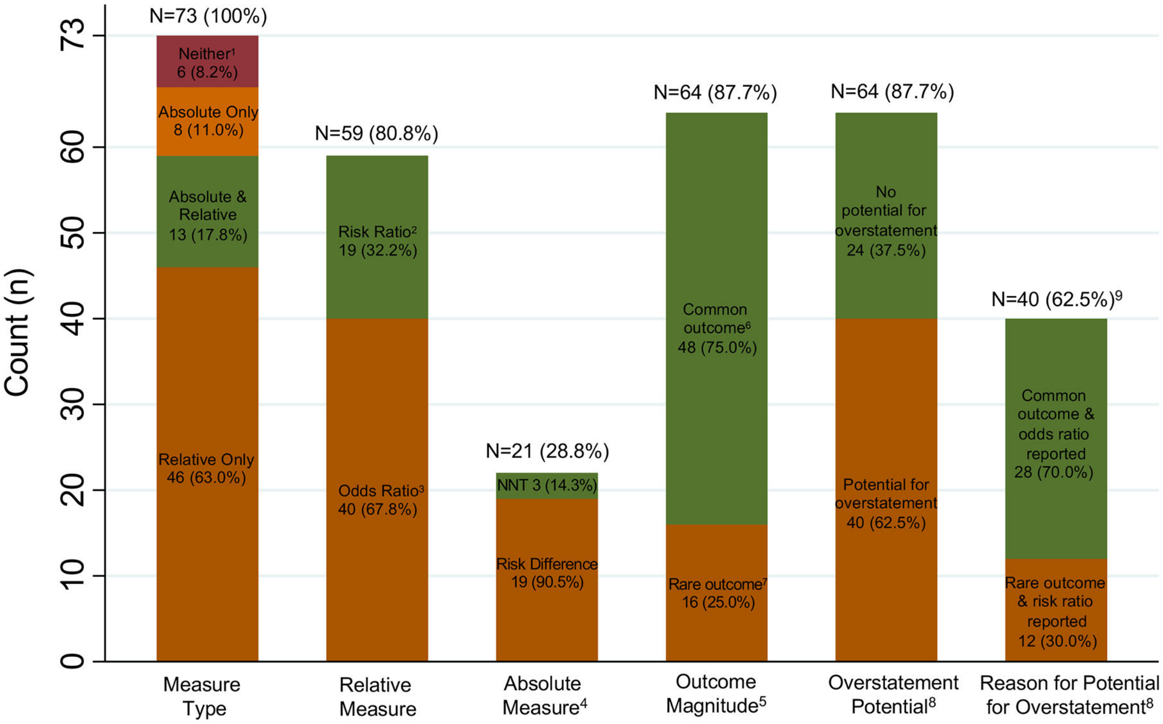

Figure 1: Summary of reporting of results for CRTs with primary binary outcome (for N=73 CRTs included in systematic review)

Abbreviation: NNT – Number Needed to Treat. Note: Percentages above each bar are out of 73, except for the final bar, which is out of 64. Percentages of segments within each bar sum to 100%, except for absolute measure type.

1. 1 reported proportions per arm with a p-value; 4 reported only proportions per arm with neither a p-value nor other measure of statistical significance; 1 reported only a p-value with no proportions by arm.
2. 2 articles reported “difference-in-differences” as the between-arm (i.e. intervention vs. control) difference in the within-arm change in proportion from baseline to endline.
3. 2 articles reported a ratio of odds ratios (ROR) in the abstract and main text. More specifically, 1 ROR was a comparison between intervention and control arms of the within-arm odds ratio for baseline to endline change, and 1 ROR was the ratio of the between-arm odds ratio (i.e. intervention effect) based on two levels of a postrandomization covariate.
4. Categories are not mutually exclusive. One paper reports both a NNT and a risk difference.
5. For 64 articles that report both an intervention effect as well as outcome proportions by arm. Note that, of the 73 articles, 6 report no intervention effect (neither absolute nor relative effect) and an additional 3 articles do not report outcome proportions by arm and therefore outcome magnitude cannot be ascertained.

6. Common outcome defined as: risk of the primary binary outcome is > 10% in both the intervention arm and the control arm.

7. Rare outcome defined as: risk of the primary binary outcome is ≤ 10% in either the intervention arm or the control arm.

8. CRTs are classified as having potential for overstatement of intervention impact if they either:

   - Report only a relative measure (i.e. with no absolute measure) when the outcome is rare, or,

   - Report an odds ratio when the outcome is common.

9. Among the 64 articles that report both an intervention effect as well as outcome proportions by arm.

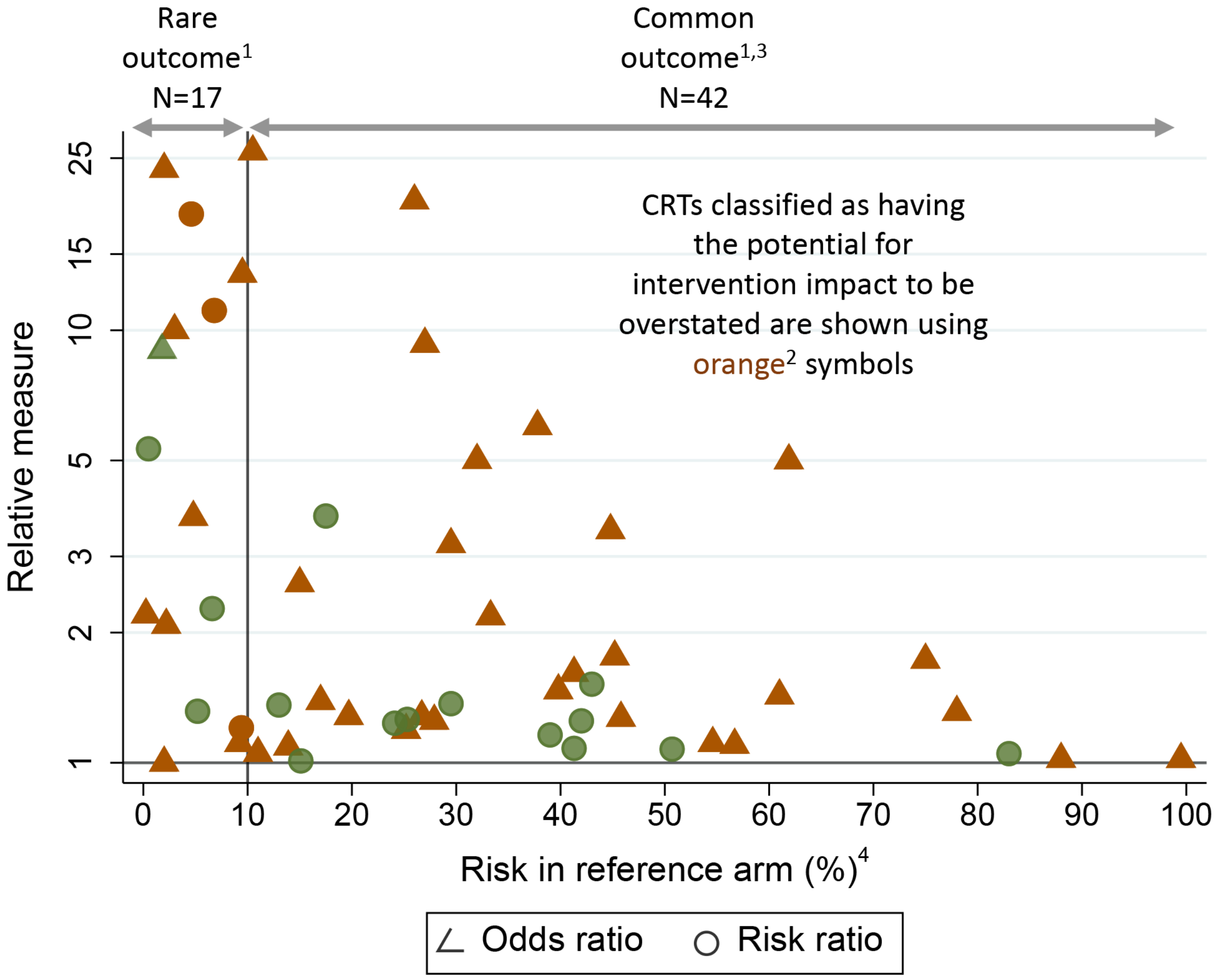

Figure 2: Magnitude of relative effect in relation to reference risk for CRTs with primary binary outcome (for N=59[1] CRTs in systematic review that reported a relative effect)

1. 55 of the 59 articles reporting relative measures are included in this figure. Four articles cannot be included for the following reasons:

   - 1 article with rare outcome that reports only a relative measure has an odds ratio of 512.

   - 1 article with rare outcome reports risk ratios, but only stratified by gender.

   - 2 articles with common outcome each report a ratio of odds ratio (ROR): 1 ROR was a ratio for intervention arm vs. control arm of the within-arm odds ratio for baseline to endline change, and 1 ROR was the ratio between two levels of a post-randomization covariate of the between-arm odds ratio (intervention vs. control) of the intervention effect.

2. CRTs classified as having the potential for intervention impact to be overstated are shown using orange symbols. These are for articles that either:

    - Report only a relative measure (i.e. with no absolute measure) when the outcome is rare (defined as risk ≤10% in either arm, or, equivalently, reference risk ≤10%), or,

    - Report an odds ratio when the outcome is common (defined as risk >10% in both arms, equivalently reference risk >10%).

3. 28 studies reported an odds ratio and had a common outcome (risk > 10% in both arms). The magnitude of potential for overstatement of impact was quantified via the ratio of the odds ratio (OR) and the risk ratio (RR). This was estimated for each CRT using the reported risks by arm to obtain both the OR and RR, from which the ratio OR/RR was calculated. Mean (SD) [min, 25th, 50th, 75th percentiles, max] ratio: 1.4 (0.6) [1, 1.06, 1.21, 1.43, 3.2].

4. For ease of visualization, the horizontal axis shows the reference risk, which is the smaller of the reported intervention-arm and control-arm risks.