The impact of using different statistical effect sizes when analysing harms in rheumatology trials:

Protocol for a meta-epidemiological study

05.12.2017

Collaborators/Authors:

Dario Kohlbrenner[1,2]/Sabrina M. Nielsen[1], Louise Klokker[1], John P. A. Ioannidis[3], Yoon K. Loke[4], Daniel E. Furst[5], Amye L. Leong[6], Maarten Boers[7], Simon Tarp[1], Peter Tugwell[8], Lee S. Simon[9], George Wells[10], and Robin Christensen[1]

1: Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
2: Department of Nursing and Allied Health Professions Education, University Hospital Zurich, Zurich, Switzerland
3: Department of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, California, USA
4: School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK
5: Division of Rheumatology, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
6: Food Science & Technology Programme, Department of Chemistry, National University of Singapore
7: Department of Epidemiology & Biostatistics; Amsterdam Rheumatology and Immunology Center; VU Medical Center, Amsterdam, Netherlands
8: Department of Medicine, University of Ottawa, Ottawa, ON, Canada
9: SDG LLC, One Mifflin Place, Suite 400, Cambridge, MA 02138, USA
10: Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada

Contact:

Robin Christensen
Senior Biostatistician, MSc, PhD
Professor of Clinical Epidemiology, adj*

Head of Musculoskeletal Statistics Unit (MSU)
The Parker Institute, Bispebjerg and Frederiksberg Hospital,
Nordre Fasanvej 57; DK-2000 Copenhagen F
Denmark

E-mail: Robin.Christensen@regionh.dk
Tel: +45 3816 4165
Fax: +45 3816 4159
Skype: dr.robin.christensen
ORCID iD 0000-0002-6600-0631
URL: http://www.parkerinst.dk/staff/robin-christensen

ABSTRACT

Background: Reporting of harm-associated outcomes remains infrequent in randomised controlled trials (RCTs). This creates difficulties in judging the Benefit-to-Harm ratio of a specific intervention. Most guidance documents imply that the intention-to-treat (ITT) principle also applies for the analyses of harms. This may lead to an underestimation of the harm, but the extent of such bias is unknown. Our objective is to explore the impact of using different effect sizes and analysis populations, i.e. the ITT population and the "as observed" population, in the analysis of harms in rheumatology trials.

Methods and analysis: Systematic reviews and meta-analyses comparing an experimental intervention with a comparator in patients with osteoarthritis and/or rheumatoid arthritis will be included in this meta-epidemiological study. For each trial, eight different effect sizes for the outcome, serious adverse events, will be calculated: risk ratio (RR) and risk difference (RD) in the ITT population (RR_{ITT} and RD_{ITT}), RR and RD in the “as observed” population (RR_{AsO} and RD_{AsO}), rate ratio and rate difference ($RR_{pyrs,e}$ and $RD_{pyrs,e}$), and rate ratio and rate difference that allow for multiple events per patient ($RR_{pyrs,e*}$ and $RD_{pyrs,e*}$). For each effect size, the Wald-z-test values will be used to compare the effect sizes in order to explore how the results may differ, as well as the effect size that apparently will make an intervention look more harmful.

Discussion: This study will provide an overall picture of how the use of different effect sizes and analysis populations in the analysis of harms may impact the results in rheumatology trials. We anticipate that these findings could provide an initial basis to recommend the optimal specific effect size and analysis population.

Keywords: analysis population, harms, meta-epidemiology, rheumatology.

INTRODUCTION

Background

It is well known that excluding randomised patients from the analyses in a randomised controlled trial (RCT) may lead to (attrition) bias.[1 2] The recommended solution is to analyse the results according to the intention-to-treat (ITT) principle.[3] According to this principle, the results of all randomised participants are included in the analysis and attributed to the treatment group to which they were originally allocated; the approach ignores protocol deviations (i.e. deviation from the intended intervention), missing data and noncompliance. Analyses based on the ITT population maintain the prognostic balance created by randomisation, but are as a result 'conservative', making it less likely to observe differences between treatment groups and which is sometimes called 'biased towards the null hypothesis (of no difference)'. This 'conservatism' creates an extra hurdle to prove added benefit of experimental interventions, the price paid for optimum validity. This is potentially critical, since RCTs contribute to meta-analyses assessing harmful effects, which are occasionally highly controversial and are likely to impact future clinical practice.[4 5]

Adequate reporting of harm associated outcomes remains infrequent in RCTs.[6 7] Saini et al. showed that 76% of the studies included in reviews did not adequately report the primary harm outcome, hampering assessment of the Benefit to Harm ratio.[6] As emphasised by the Consolidated Standards of Reporting Trials (CONSORT), reporting of harms should be considered an essential part of every study; the reporting should contain the absolute numbers (including explicit numerators and denominators) corresponding to the analyses of harms and also provide the absolute risk per arm and per adverse event type, grade and seriousness.[3]

Even though RCTs increasingly try to consider all outcomes (both beneficial and harmful), data on adverse events (AEs) may be more fragmented and incomplete, and given more cursory treatment than efficacy data.[7] Although there is no empirical evidence for harm outcomes, most guidance documents imply that the ITT principles also apply for the corresponding analyses of harms. However, as ITT procedures are 'conservative' they also increase the likelihood of finding no difference in harms, potentially an underestimation of harm in the experimental group; the extent of such bias is unknown.[8]

Important AEs are rarely (if ever) the primary outcome of included studies. As a consequence of RCTs being the core part of the pivotal (drug approval) pathway, statistical methods used to report these trial findings might need to differ from those that only address efficacy/effectiveness. One such aspect is the impact of different analysis populations on safety data. Alternatives to the ITT approach are to report on the basis of "as-observed" or to take the time under observation (or ideally under exposure) into account by using "Events-per-patient years". If standard analysis practice is to be changed robust empirical evidence is needed. For this purpose, meta-epidemiological research can be carried out.[9] To our knowledge there is no empirical evidence on the impact of using different analysis populations, such as the intention-to-treat population compared to an "as observed" population, as well as using time-under-observation in the analysis of harms, and thus, no guidance exists which effect sizes to rely on for harms in RCTs.

Objective

We will explore the impact of using different effect sizes and analysis populations, i.e. the intention-to-treat population compared to an "as observed" population, as well as the use of time-under-observation, in the analysis of harms in randomised trials, by using the database from the Cochrane Musculoskeletal Group (CMSG) of systematic reviews of controlled trials in Rheumatoid Arthritis (RA) and Osteoarthritis (OA).

METHODS

Protocol and registration

This protocol has been registered on PROSPERO (CRD42017074135) and conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines for reporting systematic reviews and meta-analyses.

Eligibility criteria

Systematic reviews and meta-analyses (SRMA) analysing RCTs published in the Cochrane database of systematic reviews (CDSR) will be used to conduct this study. Only SRMA focusing on treatments in RA and/or OA and that compare an experimental intervention with a comparator

(e.g., sham, placebo, standard practice or no intervention control) will be considered eligible for inclusion. Primary studies will be considered for eligibility if they are included as a randomised trial in a SRMA and a full text in English is available.

Information sources, search strategy, and selection process

We will use the latest version of the CDSR (http://www.cochranelibrary.com/) using the following terms:

- osteoarthritis OR rheumatoid

Additionally we will use a pragmatic search strategy via PubMed using (and combining) the following terms:

- "The Cochrane database of systematic reviews"[Journal]

AND

- (osteoarthritis[tiab] OR rheumatoid[tiab])

All SRMAs, and subsequently trial reports/manuscripts will be independently evaluated for eligibility by two reviewers (DK and SMN). Any disagreement will be resolved by discussion and, if necessary, by involvement of a third reviewer (RC).

Data collection process

For each eligible SRMA, we will extract the CDSR-registration number, name of the first author, publication year, condition (i.e. RA, OA or mixed), and title.

Each eligible trial will be assigned a trial ID, and the following will be extracted for the full trial: name of the first author, acronym (if available), publication year, condition (i.e. RA, OA or mixed), initiation (i.e. investigator or industry initiated), type of intervention, total number of patients randomised, time until latest follow-up. For each group in the trial, the number of patients randomised to each group, the number of patients at follow-up, the total-person-years, number of patients experiencing serious adverse events (SAEs), and total number of SAEs will be extracted. For outcomes collected at different times, the latest follow-up will take preference.

If total-person-years are not available in each group, it will be estimated by assuming a linear dropout rate between baseline and end of controlled period (i.e. the average area under the curve).[10]

Effect sizes

Eight different effect sizes, the risk ratios for ITT and as observed (RR_{ITT} and RR_{AsO}), the rate ratios for number of patients or number of events ($RR_{pyrs,e}$ and $RR_{pyrs,e*}$), and similar risk/rate differences (RD_{ITT}, RD_{AsO} $RD_{pyrs,e}$ and $RD_{pyrs,e*}$) will be used (**Table 1**).

Table 1: Effect sizes with corresponding variances

Population	Risk ratio/ rate ratio	Var(ln[RR])	Risk difference/ rate difference	Var(ln[RD])
ITT	$\frac{e_I/N_I}{e_C/N_C}$	$\frac{1}{e_I}+\frac{1}{e_C}-\frac{1}{N_I}-\frac{1}{N_C}$	$\frac{e_I}{N_I}-\frac{e_C}{N_C}$	$\frac{e_I\cdot(N_I-e_I)}{N_I^3}+\frac{e_C\cdot(N_C-e_C)}{N_C^3}$
AsO	$\frac{e_I/n_I^*}{e_C/n_C^*}$	$\frac{1}{e_I}+\frac{1}{e_C}-\frac{1}{n_I*}-\frac{1}{n_C*}$	$\frac{e_I}{n_I*}-\frac{e_C}{n_{C*}}$	$\frac{e_I\cdot(n_I*-e_I)}{n_I^{*3}}+\frac{e_C\cdot(n_C*-e_C)}{n_C^{*3}}$
Pyrs,e[a)]	$\frac{e_I/pyrs_I}{e_C/pyrs_C}$	$\frac{1}{e_I}+\frac{1}{e_C}$	$\frac{e_I}{pyrs_I}-\frac{e_C}{pyrs_C}$	$\frac{e_I}{pyrs_I^2}+\frac{e_c}{pyrs_c^2}$
Pyrs,e*[a)]	$\frac{e_I*/pyrs_I}{e_C*/pyrs_C}$	$\frac{1}{e_I*}+\frac{1}{e_C*}$	$\frac{e_I*}{pyrs_I}-\frac{e_C*}{pyrs_C}$	$\frac{e_I*}{pyrs_I^2}+\frac{e_c*}{pyrs_c^2}$

AsO, as observed population; C, comparator group; e, number of patients with events; e*, total number of events; I, intervention group; ITT, intention-to-treat population; N, number of patients randomise; n*, number of patients with information on the outcome; n, number of patients at latest follow-up; pyrs, patient-years under observation (serving as a proxy for patient-years under exposure); T, duration of the trial in years; var, variance.

[a)] For each randomized comparison, patient-years will be calculated as $pyrs_I = ½T(N_I + n_I)$ and $pyrs_c = ½T(N_c + n_c)$ unless directly reported in the publication.

The outcome will be patients experiencing SAEs (or total number of SAEs, i.e. allowing more events per patient, when calculating $RR_{pyrs,e*}$ and $RD_{pyrs,e*}$). In case of zero-cells we will use a continuity correction based on the reciprocal of the group (i.e. treatment or control) size opposite the zero cell.[11] Trials with multiple groups will be treated like individual trials, referred to as "randomized comparisons" (i.e. a trial with three groups, whereas two are active interventions will generate two randomized comparisons with the comparator group). For each effect size, the Wald-z-test value will be calculated as z(measure) = Signal/noise, where the measure can be any of the effect sizes, and SE is the standard error (for the relative measures, these are analysed on the log scale) [12].

Evidence synthesis

Overall, the evidence synthesis will consist of exploring how the results of the effect sizes differ, as well as the effect sizes that apparently will make an intervention look more harmful. For this purpose, for all eligible randomised comparisons, the z-value for the different effect sizes will be plotted against each other according to all the protocolised pairwise combinations. Since a p-value can be calculated as $p = \exp(-0.717 \cdot z - 0.416 \cdot z^2)$ [12], an absolute z-value above 1.96 or below -3.68 corresponds to a "statistically significant" p-value (i.e. below 0.05). Disagreements regarding statistical significance (i.e. cases for which the choice of effect size would have affected the conclusions) will be marked on the plots (e.g. by coloured areas), as well as summarised in a table. To test if disagreements exist regarding statistical significance between pairs of effect sizes, the McNemar test for dependent samples will be applied [13].

Statistically important discrepancies, i.e. if a pair of effect sizes results in statistically significant conclusions of different directions, will be described as well.

Our overall evaluation of the effect sizes will be based on the obtained empirical evidence. If an effect size tends to give results that are so different that this may impact the conclusions of a trial, we will interpret this as an important issue that needs to be taken into account when data for harms are analysed in the future. We will furthermore try to identify which effect size apparently will make interventions look more harmful compared to the other effect sizes.

DISCUSSION

Our findings will provide an initial basis for influencing the current practice on how to report harm results in RCTs. This study delivers results both for authors of future RCT reports and clinical practitioners, regarding the generalizability of study results and will serve the safety of all patients receiving treatment (incl. RA and/or OA patients).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DK, SMN, LK and RC conceived of the study, and participated in its design and helped to draft the protocol manuscript. DK wrote the first draft of the protocol manuscript. All authors edited the protocol manuscript and read and approved the final version.

Funding

The Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-13-309).

References

1. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *International journal of epidemiology* 2005;34(1):79-87. doi: 10.1093/ije/dyh300 [published Online First: 2004/11/25]
2. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *Bmj* 2009;339:b3244. doi: 10.1136/bmj.b3244 [published Online First: 2009/09/09]
3. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of internal medicine* 2004;141(10):781-8. [published Online First: 2004/11/17]
4. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *The New England journal of medicine* 2007;356(24):2457-71. doi: 10.1056/NEJMoa072761 [published Online First: 2007/05/23]
5. Christensen R, Kristensen PK, Bartels EM, et al. Efficacy and safety of the weight-loss drug rimonabant: a meta-analysis of randomised trials. *Lancet (London, England)* 2007;370(9600):1706-13. doi: 10.1016/s0140-6736(07)61721-8 [published Online First: 2007/11/21]
6. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *Bmj* 2014;349:g6501. doi: 10.1136/bmj.g6501 [published Online First: 2014/11/25]
7. Hodkinson A, Kirkham JJ, Tudur-Smith C, et al. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ Open* 2013;3(9):e003436. doi: 10.1136/bmjopen-2013-003436 [published Online First: 2013/10/01]
8. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj* 2011;343:d5928. doi: 10.1136/bmj.d5928 [published Online First: 2011/10/20]

9. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Statistics in medicine* 2002;21(11):1513-24. doi: 10.1002/sim.1184 [published Online First: 2002/07/12]
10. Tarp S, Eric Furst D, Boers M, et al. Risk of serious adverse effects of biological and targeted drugs in patients with rheumatoid arthritis: a systematic review meta-analysis. *Rheumatology (Oxford, England)* 2016 doi: 10.1093/rheumatology/kew442 [published Online First: 2016/12/26]
11. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in medicine* 2004;23(9):1351-75. doi: 10.1002/sim.1761 [published Online First: 2004/04/30]
12. Altman DG, Bland JM. How to obtain the P value from a confidence interval. *BMJ* 2011;343 doi: 10.1136/bmj.d2304
13. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;12(2):153-7. [published Online First: 1947/06/01]