PROTOCOL

Studies contributing and not contributing individual participant data (IPD) in IPD meta-analyses

Yasushi Tsujimoto[1], Tomoko Fujii[2], Akira Onishi[3], Kenji Omae[1], Yan Luo[3], Hissei Imai[3], Sei Takahashi[1], Takahiro Itaya[1], Claire Pinson[4], Toshi A Furukawa[3]

1. Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine
2. Department of Epidemiology and Preventive Medicine, Kyoto University Graduate School of Medicine
3. Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine
4. Department of Cognitive Neuroscience and Evolutionary Psychology, Harvard Colledge

Introduction

Systematic reviews (SRs) with individual patient data (IPD) meta-analysis are considered to provide more statistical power and consistent analyses as well as opportunities for more valid subgroup analyses, in comparison with SRs based on aggregate data (AD) extracted from published trial reports [1-3]. Encouragement to share IPD from randomized controlled trials (RCTs) has risen in the scientific literature, and the number of SRs with IPD meta-analysis has been increasing drastically over the past few years [4-9].

However, pictures are not all that rosy with SRs with IPD meta-analyses. First, they require increased time and efforts on the part of the review authors in collecting data, namely in requesting IPD data [1, 10, 11]. Second, in part due to this first difficulty, SRs with IPD meta-analysis have a risk of data availability bias when all IPD data requested could not be procured [2, 10, 12, 13]. If unavailability of IPD is associated with the direction or size of the intervention effect, studies that are available for IPD analysis will not be representative of the whole evidence, and the results of such IPD meta-analysis may be misleading. A previous review reported that only 25% and 43% of the published IPD meta-analyses retrieved 100% and 80% or more, respectively, of IPD from the relevant trials [10]. Additionally, more than half of IPD meta-analyses without 100% retrieval rate did not report specific reasons why some IPD were not

available. The review also revealed that several SR-level factors such as authorship policy and the inclusion of only RCTs in the review were associated with higher data availability [10].

However, to date, data availability bias has been discussed only anecdotally, narratively or theoretically and there has been no systematic examination that quantified this bias [2, 6, 10, 13, 14]. If there truly exists data availability bias, we should find a difference in the meta-analytic results between studies contributing and not contributing IPD. Moreover, RCT-level factors associated with data availability are still unknown. The purposes of this study are therefore two-fold: (i) to assess RCT-level factors associated with provision of IPD data, and (ii) to quantify data availability bias among SRs with IPD meta-analysis with less than 100% retrieval rate.

Methods

Eligibility criteria

All therapeutic RCTs included in SRs that fulfill the following criteria will be eligible: (i) SRs with IPD meta-analysis, (ii) SRs that include only RCTs comparing an active intervention against a control condition in terms of a dichotomous outcome, (iii) SRs that report full reference list of the included RCTs, and (iv) SRs written in English. The following SRs will be excluded: (v) SRs published before 2011, (vi) SRs where all included RCTs provide IPD data, (vii) SRs of diagnostic or prognostic studies, and (vii) SRs with network meta-analysis.

Search methods

We will consult the reference lists from the recent comprehensive review of IPD meta-analyses conducted by Nevitt et al [10]. We will also search MEDLINE via Ovid using the search strategy of the above review to identify relevant SRs after their search date [10].

Study selection

Four researchers (YT, TF, KO, and AO) will screen the titles and abstracts of articles identified by the MEDLINE search. We will pool the potentially eligible SRs and reference lists from the review conducted by Nevitt et al [10]. We will assess the eligibility based on a full-text review. This study will therefore include the RCTs included in the eligible SRs.

Data extraction

Six researchers (YT, TF, KO, AO, YL and CP) will independently extract the following RCT-level data from the included RCTs; year of publication, sample size, allocation concealment,

industrial sponsorship, publication status (full-publication or not), data sharing statement (available, unavailable, or unclear), journal impact factor, country of origin and language.

We will extract the following SR-level data from the included SRs; year of publication, the number of included RCTs, types of review (pharmacological or non-pharmacological, adult or pediatric, and Cochrane or non-Cochrane), and funding.. We will screen all outcomes with IPD meta-analysis.

As the SR may provide several meta-analytic results for the same class of interventions and comparators, we will select one combination of comparison and outcome per SR using the following decision rule: (1) Comparison of an experimental intervention versus control using risk ratio (RR) or odds ratio (OR), (2) Not composite outcome, (3) Not comparison of two active treatments or meta-analyses of adverse events, or meta-analysis for subgroup analysis, and (4) If several outcomes are eligible per review, we will select the one with the largest number of trials, and if they include the same number of trials, we will select the one described first.

For the selected combination of comparison and outcome, we will extract the number of events and participants in the intervention and control groups, respectively, from the SR. If they are not available, we will go back to the original RCTs and extract them. We will also extract the RR or OR from the reported IPD meta-analysis of RCTs.

Statistical analysis

We will describe RCT features classified by the provision of IPD or not. We will then explore the factors associated with the provision of IPD using logistic regression and student's T-test for dichotomous and continuous variable, respectively. We will examine the association between RCT factors and provision of IPD using mixed effects logistic regression model with fixed factors (year of publication, sample size, adequate allocation concealment, industrial sponsorship, publication status (full-publication or not), data sharing statement (available or not), journal impact factor, and language (written in English or not)) and a random intercept for the SR to account for the clustering effects within the SR. We select adequate allocation concealment as a marker of study quality because the feasibility of blinding and its impact on the outcome depend on each research question.

Our primary outcome is a discrepancy of AD meta-analytic results between RCTs contributing IPD and RCTs not contributing to IPD. Secondary outcome includes the discrepancy between IPD meta-analytic results of RCTs contributing IPD and AD meta-analytic results of RCTs not contributing to IPD.

We will summarize the characteristics of SRs. For the primary outcome, we will calculate each ratio of odds ratios (ROR) between AD meta-analyses of RCTs contributing to IPD and AD meta-analysis of RCTs not contributing to IPD. If the number of events or participants of RCTs contributing to IPD is missing, we will use IPD meta-analytic results instead. Each meta-analytic result will be re-coded so that an OR <1 favors the intervention arm. We will then combine ROR using the following two-step approach proposed by Sterne et al [15]. First, for each SR, we will estimate a ROR by using a random-effects meta-regression. A ROR <1 indicates larger treatment effect estimate for AD meta-analysis of RCTs contributing to IPD than RCTs not contributing to IPD. We will then estimate combined ROR across SRs and the 95 % CI by using random-effects meta-analysis model. The heterogeneity between SRs will be quantified with the I2 statistic.

For discrepancy between IPD meta-analytic results of RCTs contributing to IPD and AD meta-analytic results of RCTs not contributing to IPD, we will repeat the same analysis for the primary outcome.

For the primary outcome, we will perform subgroup analysis using univariable meta-regression model adding SR-level data (year of publication, the number of included studies, types of review, or funding) as covariates. For sensitivity analysis, we will exclude RCTs for which we impute the results. We do not plan any sensitivity analysis to adjust RCT-level factors on ROR because RCT-level factors are usually fixed before the provision of IPD.

Continuous variables will be expressed as mean (standard deviation) and categorical variables will be shown as numbers with the percentage. A two-sided p value smaller than 0.05 will be considered as a statistically significant difference. We will use Stata/SE, V.14.0 (StataCorp, College Station, TexasX, USA) for all analyses.

Reference:

1. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual

patient data: is there a difference? Lancet. 1993;341:418-22.
2. Tudur Smith C, Marcucci M, Nolan SJ, Iorio A, Sudell M, Riley R, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. Cochrane Database Syst Rev. 2016;9:MR000007.
3. Rogozinska E, Marlin N, Thangaratinam S, Khan KS, Zamora J. Meta-analysis using individual participant data from randomised trials: opportunities and limitations created by access to raw data. Evidence-based medicine. 2017;22(5):157-62.
4. Rathi V, Dzara K, Gross CP, Hrynaszkiewicz I, Joffe S, Krumholz HM, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ. 2012;345:e7570.
5. Tudur Smith C, Nevitt S, Appelbe D, Appleton R, Dixon P, Harrison J, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. Trials. 2017;18(1):319.
6. Huang Y, Mao C, Yuan J, Yang Z, Di M, Tam WW, et al. Distribution and epidemiological characteristics of published individual patient data meta-analyses. PLoS One. 2014;9(6):e100151.
7. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. JAMA. 2013;309(13):1355-6.
8. Data CoSfRSoCT. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.: Washington (DC): National Academies Press (US); 2015 Apr 20.
9. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing Clinical Trial Data--A Proposal from the International Committee of Medical Journal Editors. N Engl J Med. 2016;374(4):384-6.
10. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ. 2017;357:j1390.
11. Higgins JP, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011.
12. Huang Y, Tang J, Tam WW, Mao C, Yuan J, Di M, et al. Comparing the Overall Result and Interaction in Aggregate Data Meta-Analysis and Individual Patient Data Meta-Analysis. Medicine (Baltimore). 2016;95(14):e3312.
13. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ. 2012;344:d7762.

14. Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. J Clin Epidemiol. 2007;60(5):431-9.
15. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11):1513-24.