Sex-based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses

Joshua D Wallach, Patrick G Sullivan, John F Trepanowski, Ewout W Steyerberg, John PA Ioannidis

Department of Health Research and Policy, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, USA,
Joshua D Wallach, Doctoral Student in Epidemiology and Clinical Research

Departments of Health Research and Policy and Medicine, and Meta-Research Innovation Center at Stanford (METRICS), Stanford, CA, USA,
Patrick G Sullivan, Resident Physician

Stanford Prevention Research Center, Stanford University, Stanford, California, USA
John F Trepanowski, Postdoctoral Fellow

Department of Public Health, Erasmus MC, Rotterdam, the Netherlands
Ewout W Steyerberg, Professor

Departments of Medicine, Health Research and Policy, and Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, USA
John PA Ioannidis, Professor

Background

Subgroup analyses in randomized controlled trials (RCTs) are commonly used to determine whether treatment effects vary across certain patient characteristics, e.g. whether the relative effect is larger among men than women.[1] Results from these analyses have been used to tailor patient care ("stratified medicine"). Male and female subgroups are frequently compared regarding their response to a medication, reflecting our knowledge of the differences in physiology, pharmacokinetics, and pharmacodynamics that exist between the sexes.[2-5] There is some evidence that women may respond differently from men to many medications and may have more adverse drug events.[2-8] For these reasons, identification and validation of sex-based subgroup differences are necessary to ensure that males and females receive optimal medical treatment.

While little is currently known about sex-based subgroup differences specifically, much evidence suggests that published claims for subgroup differences in general are frequently invalid. Trials often perform multiple subgroup analyses without correcting for multiple testing, thereby increasing the probability of false positive claims of subgroup differences.[9-11] Furthermore, subgroup analyses are often not pre-specified, which can further promote the reporting of spurious findings.[9 12 13] Even when RCTs pre-specify a subgroup analysis plan, they deviate from this protocol over 90% of the time.[14] Only about half of RCTs with subgroup analyses use the appropriate statistical test for making inferences about subgroup differences (test for interaction).[9 13 15]

Meta-analyses (MAs) published in The *Cochrane Database of Systematic Reviews (CDSR)* can be used to understand the extent to which gender-related subgroup differences occur in RCTs and whether they are validated or not by other trials on the same question. Using an MA framework, both the effect size and the precision of an RCT's subgroup findings can be evaluated for consistency across multiple RCTs.

Aims

Aim 1: To determine how often nominally statistically significant ($p<0.05$) sex-based subgroup differences (sex-treatment interactions) are seen in the MAs published in The *CDSR.*

Aim 2: To determine how often a nominally statistically significant ($p<0.05$) sex-based subgroup difference (sex-treatment interaction) is seen in the first published (oldest) RCT included in an MA, in any RCT included in a MA, and in the summary effects combining all RCTs in a MA.

Aim 3: To determine how often a nominally statistically significant ($p<0.05$) sex-based subgroup difference (sex-treatment interaction) seen in the first RCT or in any single RCT is corroborated and/or validated by the summary of data from all other RCTs in the same MA.

Data sources and Methods

We will use data from the *CDSR.* Specifically, an automated code will screen the "Data collection and Analysis" section of all reviews. The search terms will be ("gender" OR "sex" OR "men" OR "women" OR "female") AND "subgroup." These search terms were based on two initial evaluations of reviews with forest-plots that included subgroup analyses for sex/gender. All results will be screen at the title and abstract level to exclude studies that are only about one specific sex. We will then search for forest-plots that include subgroup analyses that pertain to these search terms, as well as forest plots that have any of these terms in their title, because it is possible that some reviews may present gender subgroup analyses as separate forest plots rather than sub-analyses within the same forest plot. When necessary, forest plots on men will need to be matched with the respective forest plots on women. This screening will be performed by three independent reviewers (JDW, JFT, and PGS) and the IDs of the eligible forest plots (CD number and forest plot number) will be recorded. Potential discrepancies will be arbitrated by a third reviewer (JPAI).

The following characteristics will be extracted manually by one data extractor (JDW) on all eligible forest plots that have passed both stages of screening and will be reviewed by two additional independent reviewers (JFT, PGS): condition/disease, compared treatment interventions, outcome, number of RCTs on men, number of RCTs on women, number of RCTs

with data on both men and women, total sample size (on men, women, and both). When multiple outcomes on the same comparison and disease are eligible and provide data to evaluate sex-based differences, data on all of those outcomes will be recorded separately.

For each topic, all studies that have separate data on both men and women will be considered for further quantitative analyses. RevMan (version 5.4) will be used to evaluate the sex-treatment interaction will for all comparison-outcomes from the eligible studies. The sex-treatment interaction will first be evaluated replicating the analyses that the authors originally performed. For consistency, studies with metrics based on binary data will then be standardized to risk ratios and studies with continuous outcomes will be transformed into mean differences. Hazard ratios and rate ratios will not be transformed. Each sex-treatment interaction will then be re-evaluated with the inverse-variance random effects method. After recording the total number of overlapping trials for the sex subgroups, the sex treatment interaction will be calculated using only comparison-outcomes with overlapping trials in the sex subgroup levels. As a further sensitivity analysis, the sex-treatment interaction will be re-evaluated with a fixed-effects analysis. All comparison-outcomes with a statistically significant *P*-value from the test for subgroup differences ($P < 0.05$) will be retested with a more appropriate random-effects model. In particular, we will the metafor package in R to perform a Hartung-Knapp-Sidik-Jonkman (HKSJ) method for random effects meta-analysis, which has been show to outperform the standard DerSimonian-Laird method, for all entries with statistically significant p-value for subgroup difference.

References

1. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med 2007;**357**(21):2189-94.
2. Cotreau MM, von Moltke LL, Greenblatt DJ. The influence of age and sex on the clearance of cytochrome P450 3A substrates. Clin Pharmacokinet 2005;**44**(1):33-60.
3. Schwartz JB. The influence of sex on pharmacokinetics. Clin Pharmacokinet 2003;**42**(2):107-21.
4. Legato MJ. Gender and the heart: sex-specific differences in normal anatomy and physiology. J Gend Specif Med 2000;**3**(7):15-8.
5. Kashuba AD, Nafziger AN. Physiological changes during the menstrual cycle and their effects on the pharmacokinetics and pharmacodynamics of drugs. Clin Pharmacokinet 1998;**34**(3):203-18.
6. Xhyheri B, Bugiardini R. Diagnosis and treatment of heart disease: are women different from men? Prog Cardiovasc Dis 2010;**53**(3):227-36.
7. Makkar RR, Fromm BS, Steinman RT, et al. Female gender as a risk factor for torsades de pointes associated with cardiovascular drugs. Jama 1993;**270**(21):2590-7.
8. Drici MD, Clement N. Is gender a risk factor for adverse drug reactions? The example of drug-induced long QT syndrome. Drug Saf 2001;**24**(8):575-85.
9. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987;**317**(7):426-32.

10. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;**355**(9209):1064-9.
11. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. N Engl J Med 2006;**354**(16):1667-9.
12. Sun X, Briel M, Busse JW, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. Bmj 2011;**342**:d1569.
13. Pocock SJ, Assmann SE, Enos LE, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;**21**(19):2917-30.
14. Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. Bmj 2014;**349**:g4539.
15. Pocock S. *Clinical Trials : A Practical Approach*. 1 ed: Wiley, 1983.