BMJ

BMJ 2012;344:e2022 doi: 10.1136/bmj.e2022 (Published 15 March 2012)

EDITORIALS

Subgroup analyses

The devil is in the interpretation

Andrew D Oxman *researcher*

Global Health Unit, Norwegian Knowledge Centre for the Health Services, N-0130 Oslo, Norway

Sun and colleagues found, in the linked systematic review (doi:10.1136/bmj.e1553), that about a third of a representative sample of recent randomised trials published in core clinical journals report subgroup analyses.[1] After judging these reports of subgroup analyses using 10 carefully developed predefined criteria, the authors conclude that only in very few instances can we be confident that subgroup analyses provide a better estimate of effect than the overall results of trials, and they describe in detail the reasons why. These findings are consistent with what could be expected on the basis of previous reviews and the play of chance.[2-7]

Previous reviews of published trials and protocols have found that subgroup analyses are commonly reported (38-87% of the time),[2-7] and that appropriate statistical analyses (tests of interaction) are not used 38-91% of the time.[2-5] In addition, planned subgroup analyses are commonly not reported (48-69% of the time) and 43-91% of randomised trials report subgroup analyses that were not planned.[6 7] When subgroup analyses are reported, authors claim differences in 25-69% of cases, and these results are commonly featured prominently (15-45% of the time).[2-5]

The extent to which the consistency of subgroup findings across trials can be assessed is limited by authors' failure to interpret their results in the context of a systematic review of other trials.[8]

In 1983 the authors of a paper that presented 146 subgroup analyses of the Beta Blocker Heart Attack trial, found, unsurprisingly, that the results were normally distributed. Roughly 2.5% of the subgroup analyses had results that were "significantly" worse and 2.5% had results that were "significantly" better.[9] Five years later the International Study of Infarct Survival 2 (ISIS-2) trial found that aspirin reduced mortality after heart attack overall (P<0.00001) but increased mortality by a small amount in patients born under Gemini and Libra astrological signs.[10] Six years after that, the DICE (Don't Ignore Chance Effects) collaborators in their meta-analysis of trials of DICE therapy (rolling dice) for acute stroke found that red dice are deadly, on the basis of a predefined subgroup analysis by colour of dice.[11] All of these findings illustrate the important message that chance influences the results of clinical trials and systematic reviews of trials. Unfortunately, both researchers and clinicians can easily be misled by the play of chance.

Sun and colleagues' findings provide a clear indication of the extent to which subgroup analyses are undertaken, reported, and interpreted uncritically. This has important implications for researchers, authors of systematic reviews, editors, clinicians, and patients.

There are many compelling reasons for performing subgroup analyses, but the interpretation of their findings can be challenging. Authors of trials and systematic reviews can help by limiting the number of subgroup analyses that are conducted to those with a clear prespecified rationale. They should report the rationale for conducting each subgroup analysis and specify planned subgroup analyses in study protocols, including the predicted direction of the difference. Appropriate statistical analyses should be used.

Editors can improve the situation by requiring that authors report and interpret subgroup analyses appropriately in trials and systematic reviews. Requirements should include clear and comprehensive reporting of all subgroup analyses and the extent to which criteria for evaluating the credibility of each subgroup analysis were met, and the use of language that reflects the extent to which such criteria were met. The box provides examples of appropriate language for reporting.

Clinicians and patients should, as a rule, base decisions about treatments on systematic reviews of trials and not on single trials, unless no other relevant trials exist.[8] Decisions based on poorly interpreted subgroup analyses might result in effective treatments being withheld from some patients who would benefit.[12] They might also lead to the use of ineffective or harmful treatments in some patients. In the absence of a critically interpreted subgroup analysis for which a high degree of confidence is warranted, the best estimate of effect for a subgroup is the overall effect.

Sun and colleagues' 10 criteria are useful for assessing how much confidence to place in the results of subgroup analyses and when to base a decision on a subgroup analysis rather than on the overall results. To save time, however, a simple rule of thumb could be to first ask: "Are the results of the subgroup analysis and the overall analysis different enough that they

oxman@online.no

Examples of plain language that reflects how much confidence can be placed in subgroup analyses

Very low confidence: If important criteria are not met (for example, inconsistent subgroup effects across trials that have no compelling explanation or a high probability that the apparent subgroup effect occurred by chance) report the difference in effects as a hypothesis that warrants further investigation and do not include it in the abstract or conclusions ("The difference in effect is uncertain")

Low confidence: If differences in effects probably did not occur by chance, but the estimated subgroup effect warrants low confidence because other criteria were not met, report the subgroup effect as a hypothesis and do not include it in the abstract or conclusions ("There may be a difference in effect")

Moderate confidence: If most of the criteria are met and there probably is an important subgroup effect, report it as probable ("There probably is a difference in effect")

High confidence: If all or nearly all of the criteria are met and a high degree of confidence is warranted, report it without qualification ("There is a difference in effect")

would lead to different decisions?" If the answer is no, the detailed criteria do not need to be applied.

Competing interests: The author has completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declares: no support from any organisation for the submitted work; 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.

Provenance and peer review: Commissioned; not externally peer reviewed.

1 Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;344:e1553.
2 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;21:2917-30.
3 Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JDF, Maas AIR. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery* 2005;57:1244-53.
4 Aulakh AK, Anand SS. Sex and gender subgroup analyses of randomized trials. *Womens Health Issues* 2007;17:342-50.
5 Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;357:2189-94.
6 Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
7 Boonacker CWB, Hoes AW, van Liere-Visser K, Schilder AGM, Rovers MM. A comparison of subgroup analyses in grant applications and publications. *Am J Epidemiol* 2011;174:219-25.
8 Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet* 2010;376:20-1.
9 Furberg C, Byington R. What do subgroup analyses reveal about differential response to beta-blocker therapy? The Beta-Blocker Heart Attack trial experience. *Circulation* 1983;67:98-101.
10 ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;2:349-60.
11 Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994;309:1677-81.
12 Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.

Cite this as: *BMJ* 2012;344:e2022