Received: 5 April 2017 | Revised: 1 November 2017 | Accepted: 30 November 2017

DOI: 10.1002/sim.7595

LETTER TO THE EDITOR

WILEY Statistics in Medicine

New research strategy with ambiguous implications: A comment on "Planning future studies based on the conditional power of a meta-analysis"

Roloff et al provide conditional power formulas for a future meta-analysis based on an already existing meta-analysis judged to be inconclusive and use them to determine sample sizes and a number of additional clinical trials to arrive at a conclusive updated meta-analysis including all studies.[1] In the following, we discuss the implications of this meta-analysis-based research strategy in comparison with a stand-alone study-based research strategy. The calculations by Roloff et al are done under the assumption of a fixed-effects model (FEM) and random-effects meta-analysis model (REM). The main difference between both models is the assumption of heterogeneity, and therefore, whether there is a common underlying effect size θ or whether study effects are following a distribution with mean θ. Hence, inference is focused on the common effect size under a FEM and on the mean of the distribution of all (heterogeneous) effect sizes under a REM. In all following formulas, the variance estimates of i already observed studies and k new studies and the heterogeneity are σ_i^2, σ_{new}^2, and τ^2, respectively. The study-effect estimates of the former studies and new studies are defined as $Y_{\text{old},\,i}$ and Y_{new}.

First, we consider the special case, that only 1 new study will be conducted, which is planned to detect a prespecified effect with a power of 80% and therefore provides stand-alone evidence against the null hypothesis (stand-alone study). If the meta-analysis is updated with this new trial by using the FEM, the treatment effect will be estimated as

$$\widehat{\theta}_{\text{FEM}} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} Y_{\text{old},i} + \frac{1}{\sigma_{\text{new}}^2} Y_{\text{new}}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} + \frac{1}{\sigma_{\text{new}}^2}} \sim N\left(\theta, \frac{1}{\sum_{i}^{n} \frac{1}{\sigma_i^2} + \frac{1}{\sigma_{\text{new}}^2}}\right) \tag{1}$$

The assumption of a REM and equal heterogeneity in the old and new meta-analysis leads to the treatment effect estimate

$$\widehat{\theta}_{\text{REM}} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + \tau^2} Y_{\text{old},i} + \frac{1}{\sigma_{\text{new}}^2 + \tau^2} Y_{\text{new}}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2 + \tau^2} + \frac{1}{\sigma_{\text{new}}^2 + \tau^2}} \sim N\left(\theta, \frac{1}{\sum_{i}^{n} \frac{1}{\sigma_i^2 + \tau^2} + \frac{1}{\sigma_{\text{new}}^2 + \tau^2}}\right) \tag{2}$$

Roloff et al calculate the power of the updated meta-analysis after including a fixed number of equally sized studies, which do not have to be planned (powered) to be conclusive on their own. [1] This suggests the segmentation of 1 stand-alone study with variance σ_{new}^2 into k smaller studies that each is $1/k$ times the size of the stand-alone study with variance $k\sigma_{\text{new}}^2$ that will be included in the updated meta-analysis.

In cases where a FEM is used for the analysis, it does not make a difference whether the stand-alone study is included as 1 study or beforehand split into smaller substudies. On the other hand, the treatment effect in the updated REM meta-analysis including the k smaller studies is estimated as

$$\widehat{\theta}_{\mathrm{REM}(k)} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2+\tau^2} Y_{\mathrm{old},i} + \frac{1}{\sigma_{\mathrm{new}}^2+\tau^2/k} Y_{\mathrm{new}}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2+\tau^2} + \frac{1}{\sigma_{\mathrm{new}}^2+\tau^2/k}} \sim N\left(\theta, \frac{1}{\sum_i^n \frac{1}{\sigma_i^2+\tau^2} + \frac{1}{\sigma_{\mathrm{new}}^2+\tau^2/k}}\right) \tag{3}$$

Here, the distribution of the treatment effect is dependent on k. Increasing $k > 1$ downweigths the heterogeneity parameter τ^2 and decreases the variance of the estimated treatment effect. Therefore, $k > 1$ in (3) leads to a narrower confidence interval compared with (2) and a power gain.[1] The magnitude of the power gain by variance reduction can be seen by maximal segmentation: For $\tau^2 > 0$ and N being the total sample size of 1 additional trial segmented into k trials, the variance in (3) is monotonically decreasing in k, and for k, increasing the variance in (3) converges to

$$\lim_{k \to \frac{N}{2}} \frac{1}{\sum_i^n \frac{1}{\sigma_i^2+\tau^2} + \frac{1}{\sigma_{\mathrm{new}}^2+\tau^2/k}} = \frac{1}{\sum_i^n \frac{1}{\sigma_i^2+\tau^2} + \frac{1}{\sigma_{\mathrm{new}}^2+\tau^2\frac{2}{N}}} \tag{4}$$

and heterogeneity is reduced up to a $\frac{2}{N}$ -fold.

To illustrate the conditional power approach for a REM, Roloff et al consider a systematic review of the role of preoperative chemotherapy for esophageal cancer including data from 8 studies involving 1729 patients as an example of a meta-analysis with moderate heterogeneity ($I^2 = 40.2\%$).[2] This meta-analysis reported a hazard ratio for the comparison of preoperative chemotherapy versus surgery alone of 0.88 (95% CI: 0.75 to 1.04). Roloff et al calculate the conditional power of the updated meta-analysis by using a more optimistic effect of 0.82. We chose the observed effect in the inconclusive meta-analysis of 0.88 as the best assumption for the updated meta-analysis effect estimate and calculated that at least 7 additional studies (with a total number of additional events of roughly 18,000) have to be conducted to reach a conditional power of 80% in the updated meta-analysis. In contrast, a stand-alone study with a total of 1921 events had a power of 80% to detect a hazard ratio of 0.88 at a significance level of 5%.

Taking the segmentation of a stand-alone study to the extremes is not what would be expected in reality, but it highlights the question: Is it appropriate to gain power for the updated meta-analysis by increasing the number of planned future studies while reducing the power of each of these planned future trials?

The use of study segmentation and subsequent meta-analysis as a strategy for future research raises some issues:

- At least 2 adequate and well-controlled studies each clearly demonstrating efficacy are demanded as a prerequisite for drug licensing by default.[3,4] The meta-analysis-based research strategy here opens a door for concluding that a drug should be considered efficacious in a situation where no individual study ever met its primary objective. This strategy is currently not supported in the field of drug licensing, where replicated randomized controlled trial results are considered higher-level evidence than meta-analysis results.
- Roloff et al assume that it would be possible to understand (and then replicate) the conditions under which heterogeneity has been observed in the inconclusive meta-analysis. They exemplify this situation with multiregional clinical trials, where heterogeneity between regions has been observed. From a purely evidentiary perspective in most of these trials, however, reasons for heterogeneity could be identified that allowed separate decision making in homogeneous subgroups (eg, Platelet Inhibition and Patient Outcomes and high-dose aspirin,[5,6] PASS, and glomerular filtration rate mutation[7]).
- In contrast, if reasons for heterogeneity of study results can be identified, a better strategy is to model this heterogeneity or conduct studies in respective homogeneous subgroups of the population. The gain in power that results from the application of the conditional power formulas in the REM meta-analysis leads to a purely technical reduction of heterogeneity without additional insights into the causes of heterogeneous study-specific treatment effects.

A research strategy in drug licensing based on randomized controlled stand-alone trials could be as follows: (a) Given a homogeneous meta-analysis, which shows a nonsignificant relevant treatment effect, we advocate the conduct of one additional stand-alone trial based on the observed effect. Here, the additional stand-alone trial will inevitably give the updated FEM meta-analysis a sufficient power, as well. (b) Given a heterogeneous meta-analysis, which features a nonsignificant relevant treatment effect, a logical research strategy would be to first make some attempts to better understand the potential reasons for heterogeneity (eg, by using subgroups as a means to understand who benefits at which risks) and then

conduct 1 additional study with well-defined inclusion and exclusion criteria. Additionally, evidence synthesis methods can be applied to get a wider picture and learn about the heterogeneity of effects, external validity, and generalizability.

However, the conditional power approach might be a useful tool in identifying heterogeneity that cannot be ignored at the planning stage of a future trial. Whenever an updated meta-analysis cannot reach sufficient power after the inclusion of 1 additional stand-alone study, Roloff's method could indicate substantial heterogeneity worth exploring. For decision making in drug licensing, however, "Individual clinical trials should always be designed to satisfy their objectives and [...] stand-alone studies (should not be) substituted by a meta-analysis of trials of inadequate size."[3]

ACKNOWLEDGEMENTS

This work has been funded partly by the FP7-HEALTH-2013-INNOVATION-1 project Advances in Small Trials Design for Regulatory Innovation and Excellence (ASTERIX) Grant Agreement No. 603160. We thank the editors and reviewers for useful comments to tailor the main messages.

ORCID

Kristina Weber http://orcid.org/0000-0001-7758-5885

Kristina Weber
Florian Lasch
Armin Koch
Institut für Biometrie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Kristina Weber and Florian Lasch contributed equally.

Correspondence
Kristina Weber, Institut für Biometrie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany.
Email: weber.kristina@mh-hannover.de

REFERENCES

1. Roloff V, Higgins JPT, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat Med*. 2013;32(1):11-24. https://doi.org/10.1002/sim.5524
2. Vogt K, Fenlon D, Rhodes S, Malthaner R. Preoperative chemotherapy for resectable thoracic esophageal cancer. *Cochrane Database Syst Rev* 2006;(3):CD001556. doi:10.1002/14651858.CD001556.pub2
3. EMA. Points to consider on application with 1. Meta-analyses; 2. One pivotal study. 2001:1-6. http://www.emea.eu.int/.
4. FDA. Providing clinical evidence of effectiveness for human drug and biological products. FDA Guid Ind. 1998:1-23. http://www.fda.gov/downloads/Drugs/Guidances/ucm078749.pdf.
5. James S, Åkerblom A, Cannon CP, et al. Comparison of ticagrelor, the first reversible oral P2Y12 receptor antagonist, with clopidogrel in patients with acute coronary syndromes: rationale, design, and baseline characteristics of the PLATelet inhibition and patient Outcomes (PLATO) trial. *Am Heart J*. 2009;157(4):599-605. https://doi.org/10.1016/j.ahj.2009.01.003
6. Mahaffey KW, Wojdyla DM, Carroll K, et al. Ticagrelor compared with clopidogrel by geographic region in the Platelet Inhibition and Patient Outcomes (PLATO) Trial. *Circulation*. 2011;124(5):544-554. https://doi.org/10.1161/CIRCULATIONAHA.111.047498
7. Mok TS, Wu Y-L, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*. 2009;361(10):947-957. https://doi.org/10.1056/NEJMoa0810699