

**EDITORIAL****Research without good questions is a waste****Good methods without good questions is counterproductive**

Good clinical epidemiological methodology is important, but it is a necessary, not sufficient condition for good research. Original ideas, creativity, clinical and scientific perceptiveness, extensive subject matter expertise and well-prepared questions and hypotheses are the starting point for scientific progress; methodology is the craft to harvest the resulting new knowledge. Both ingredients need to be present for research to succeed and make a difference.

Suppose that two productive research teams have each published a similar number of RCTs in the past 20 years. Both used excellent state-of-the-art methodology to test their hypotheses, for example on the effectiveness, added value, and safety of an intervention. Both teams also used the same statistical criteria regarding significance level and power. What to say then if the first team scored one positive result, that is, one of the 20 studies turned out to confirm the hypothesis, while the other team scored 10 positive results out of 20. Should not we say that the second team is better in developing hypotheses that have a reasonable chance to survive rigorous RCT-based testing? Of course, making such comparisons is complex, but if the procedures followed by both teams to avoid a biased chase for positive results are equally rigorous, and when a substantial number of positive results are later not refuted in other scientific or clinical settings (a process that will generally not take place for negative results), should one not conclude that the second team did a better job in advancing science and care, and in avoiding research waste? [1,2].

The idea that research teams that, using similar methodology, more often find positive results when testing their hypotheses than others, might be better, seems opposed to what is often heard: ‘as long as the methods used are appropriate, a negative result is not less important than a positive one, since for many reasons knowledge of no effect can be crucial.’ That can indeed be the case, but only if it is an answer to a well-prepared question and not a finding after more or less ‘flipping a coin’, as the latter will not bring added value. On the contrary, it would only create opportunities for false positive results and be a counterproductive use of

resources and inappropriate burdening of research subjects and patients. In fact, the general question ‘what are the determinants of a positive result in terms of confirming the studied hypothesis’, with the quality of the hypothesis and preliminary research, and even research team, among the key factors to address, would be a good topic for systematic review.

**The craft of preparing good hypotheses**

Since a good hypothesis is crucial, ethical research practice in the context of ‘clinical equipoise’, should focus on identifying the research questions that are positioned between (a) a priori certainty (where research would be unneeded, a waste and ethically unjustified), and (b) a blank initial situation of 100% uncertainty without well-prepared prior knowledge (where once again research would be unneeded, a waste and ethically unjustified) [3]. These considerations are equally relevant for both experimental and observational research; in the latter case, in an era of big data, the risk of false positive findings from ‘fishing expeditions’ is even higher. This, of course, taking into account the fact that in new research fields a well balanced exploration for unknown but potentially important knowledge can be relevant.

The key message therefore is that formulating good questions and hypotheses is no less a craft than designing the best methodology to investigate them [4–6]. Health research needs not only serendipity, defined as making discoveries or inventions by chance rather than intent. It also requires a professional process of generating ideas for new knowledge or techniques, starting from what is already known [7], identified knowledge gaps and innovative insights, or deliberately exploring unknown terrain. This is an often insufficiently fulfilled mission in research training, with important roles for scientifically trained clinicians.

Recognizing the importance of good hypotheses in addition to good methodology may have implications for the evaluation of publication bias [8]. If in studies with positive results the prepared hypotheses would overall be better than in ‘negative’ studies, it would be expected and justified that such studies would be more often published and cited. Moreover, reading studies that more or less flip a coin would be a waste of time.

### Subgroup analysis: hypothesis-driven or fishing?

The issue of using a well-developed hypothesis has since long been recognized as a relevant topic in dealing with subgroup analyses [9]. In this context, we distinguish three levels: a well-developed explicitly formulated hypothesis of effect modification (high level), which may yield useful input for a more focused RCT or for systematic review and meta-analysis; deliberate exploration (intermediate level), which might also lead to further study; and just ‘fishing’ (low level), which is below-standard research practice, as discussed in the previous paragraphs. In this connection, the paper by [Fan et al. \(9802\)](#) is important. They assessed the appropriateness and rationales of subgroup analyses planned in a Pubmed-based random sample of protocols of randomized trials and reported in subsequent trial publications. Subgroup analyses were specified in 19% of the protocols and reported in 21% of the trial publications. Justifications or rationales for subgroup analyses were rarely provided, subgroup analyses were not prespecified in most of the trial publications with subgroup analyses, and their reporting was often insufficient. The authors also found that more recently published trial protocols less often planned subgroup analyses than earlier ones, possibly since attention has been focused mainly on the problem of false positives. But, as they say, if fewer subgroup analyses are conducted, it will be difficult to examine the consistency in results of a subgroup analysis across different trials, and valuable clinical research data will be wasted. They recommend methodological guidance on subgroup analyses and sufficient reporting of results of all subgroup analyses conducted, rather than insisting on only a small number of planned subgroup analyses. With sufficient reporting of subgroup analyses from multiple trials problems such as false positive or false negative subgroup effects in individual trials, could be corrected with evidence accumulation.

### ‘Living’ accumulation is feasible

The potential for such accumulation is increasing, given the progress in conducting systematic reviews, meta-analyses, individual patient data meta-analysis, and living systematic reviews [10,11]. Recent developments in the feasibility of the latter were addressed in two articles. In the first, [Créquit and co-authors \(9801\)](#) assessed the feasibility of living network meta-analysis (NMA) taking the pace of evidence generation into account. In a systematic review of published NMAs they calculated the cumulative number of new trials for each NMA considering different update frequencies (4, 6, and 12 months), and evaluated the workload for an update relative to the initial NMA. They found that for these frequencies, respectively, a mean of 4%, 5%, and 11% workload per update was required, and concluded that living NMAs are manageable. In a second paper, [Lerner et al. 9794](#)

focused on developing and evaluating an algorithm for automatically screening citations when updating living NMAs from different medical domains. The investigators concluded that for updating an NMA after two years, their algorithm considerably diminished the workload required for screening, and that the number of missed citations remained low. They recommend integration of automatic screening for updating living NMAs be considered.

### Testing accepted practice

Hypothesis testing can also be focused on revisiting ‘classic methodological truths’. [McCambridge c.s. \(9779\)](#) for example evaluated the Hawthorne effect. Starting from evidence-informed doubt, they studied whether participants who are aware that a particular behavior is being studied will modify that behavior. In a large three arm online randomized trial among students answering the same lifestyle questions, one (control) group were told that they were completing a lifestyle survey, a second group heard that the focus of the survey was alcohol consumption, and a third group answered additional questions on their alcohol use. All groups were aware they would be followed up after one month. Between these groups no differences were found on overall drinking volume and frequency and amount per typical occasion. It was concluded that Hawthorne effect phenomena are likely to be unimportant in online alcohol studies, and that there is a need to better understand the contexts in which research participation effects are more likely to occur.

Accepted treatment practice can be challenged by testing non-inferiority hypotheses in assessing new treatments. Such research requires specific subject-matter related justification regarding [12] (a) the reason for using this study design, (b) the presumed benefit, (c) the substantiation of the choice for the non-inferiority margin, and (d) the choice of the standard treatment with which the new treatment is compared. However, in a systematic search of trials randomly selected from Medline, [Hong et al. \(9791\)](#) found that few trials reported all relevant information, and the methods as reported were not always appropriate. Most trials had a publicly available protocol/clinical trial registry, but only 45% of those mentioned the NI design in their protocol. None of the trials described the reason or justification for an NI design. When comparing the proportion of NI trials reporting on key components, publicly funded trials were not better than industry-funded trials. As to the latter, [Martins et al. \(9789\)](#) report a similar finding. In the context of an ongoing systematic review and network meta-analysis of desensitizing toothpaste trials, they could not confirm that industry funding was associated with a positive trial result. These findings should not distract us from the premise that best research practice requires that - obviously independent of the method of

financing - all research funding be declared, the potential effect on bias always considered, and if possible empirical analysis such as this, performed.

J. Andre Knottnerus  
Peter Tugwell

*E-mail address:* [anneke.germeraad@maastrichtuniversity.nl](mailto:anneke.germeraad@maastrichtuniversity.nl)  
(J.A. Knottnerus)

## References

- [1] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
- [2] A platform to share and exchange documentation, information, and resources to help increase the value of research and reduce waste in research. Available at <http://rewardalliance.net/guidelines/>. Accessed March 7, 2019.
- [3] Verhees RAF, Dondorp W, Thijs C, Dinant GJ, Knottnerus JA. Influenza vaccination in the elderly: is a trial on mortality ethically acceptable? *Vaccine* 2018;36(21):2991–7.
- [4] Vandenbroucke JP. Alvan Feinstein and the art of consulting: how to define a research question. *J Clin Epidemiol* 2002;55:1176–7.
- [5] Lipowski EE. Developing great research questions. *Am J Health Syst Pharm* 2008;65:1667–70.
- [6] Bragge P. Asking good clinical research questions and choosing the right study design. *Injury* 2010;41:S3–6. Epub 2010 May 13.
- [7] The 'Evidence-based research network. Available at <http://ebnetwork.org/>. Accessed March 7, 2019.
- [8] Suñé P, Suñé JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One* 2013;8:e54583. Epub 2013 Jan 30.
- [9] Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA* 2014;311:405–11.
- [10] Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review network. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol* 2017;91:23–30. Epub 2017 Sep 11.
- [11] Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic review network. Living systematic reviews: 2. Combining human and machine effort. *J Clin Epidemiol* 2017;91:31–7. Epub 2017 Sep 11.
- [12] Soonawala D, Dekkers OM. 'Non-inferiority'-studies: mogelijkheden en kanttekeningen. [Non-inferiority' trials. Tips for the critical reader. Research methodology 3]. *Ned Tijdschr Geneesk* 2012;156(19):A4665. Review, Dutch. PMID: 22571548.