

EDITORIAL

When becomes more in fact less?

The principles ‘in dubio abstine’ and ‘primum non nocere’ are as least as old as the language they are written in, and have been made very concrete in criticisms on iatrogenesis, medicalization, polypharmacy and polypragmasy. Clinical decision analysis [1], health technology assessment, and critical appraisal of clinical evidence [2] have provided useful methods to translate these principles in better health care by avoiding useless and potentially harmful intervention.

To day, these principles are also applied in research practice, as research resources are scarce, and superfluous research is not only a waste but may also push valuable research aside [3–5]. Moreover, useless research implies exposing research subjects to the burden and sometimes risk of being studied without justification. We must therefore also make progress in designing methods to recognize and avoid research studies without added value.

These issues are addressed in various contributions.

Stalpers *et al.* give an update of methods that can help clinicians to reduce diagnostic testing that has no added value and may even be harmful, and further elaborated the concepts ‘threshold approach to clinical decision making’ [6,7] and ‘indication area of a diagnostic test’ [8]. They developed a decision analytical model in which ‘testing’ and ‘no testing’ were compared, in order to identify the indication area of prior probabilities wherein use of a diagnostic test results in a net gain. They illustrate their approach, using a freely available website-based calculator with graphical representation developed for his purpose, with various clinical examples. In a second paper, Stalpers *et al.* investigate whether indication area and maximum diagnostic gain are robust measures in relation to test dependence [9], alternative physician’s heuristics, and varying patient’s utilities. Based on mathematical, decision heuristic, and utility analysis - and again supported by examples—the authors conclude that indication area and the maximum diagnostic gain are indeed robust measures of test performance.

In research decision making, a key question is whether adding a new study to the already available evidence, would change conclusions. Chevance *c.s.* investigated whether robustness of an existing meta-analysis can justify decisions on whether to conduct an additional study on the same research question. They present a contour plot assessment of meta-analysis, using the example of the potential impact of an additional study on the effect of statin use

on the risk of acute kidney injury. The group found that future single studies of common sample size will not considerably change the direction and magnitude of the currently evident effect estimates. They conclude that the illustrated contour approach can indeed guide decisions on whether to conduct additional studies on a relevant research question, and recommend that it should become a standard tool for the assessment of meta-analyses.

Another pertinent issue is whether noninferiority studies—in order to demonstrate whether a new treatment is not substantially less effective than an already available treatment - really have added value. Claims can be that the new treatment may have an advantage as to burden of treatment, side effects, and cost. Gayet-Ageron and her team explored characteristics of clinical trials that may influence the choice of the ‘noninferiority margin’ (NIM): the largest loss of effectiveness that is clinically negligible to establish noninferiority. They conducted an experimental survey among corresponding authors of randomized controlled trials indexed in Medline, asking their opinions on two hypothetical studies. The authors found that mortality as a primary outcome, low baseline risk with standard treatment, a less costly new treatment and fewer side effects were associated with choosing smaller NIMs. Interestingly, the population age group (adults vs. children) showed no effect on the choice of the NIM. They recommend that the process to determine the NIM should be explicitly based on clinically relevant factors, such as identified in their survey.

Although Gayet-Ageron did not find population age group to play a role in choosing the NIM, equivalence of effects of interventions among children and adults cannot a priori be assumed [10]. Janiaud *et al.* evaluated whether the therapeutic benefit in placebo controlled randomized clinical trials is different for children and adults. Based on a review of meta-analyses of trials with separate results for adults and children, for 14 of 124 drugs they found different treatment benefits for trials in adults compared with trials in children. The authors recommend that data on dose adjustment and age groups should be better reported to investigate extrapolation from adults to children.

Speaking about extrapolation, another question is whether systematic reviews of clinical trials consider applicability of study results for primary care. This was explored by Missiou and Tatsioni. In a sample from the Cochrane Database for Systematic Reviews over the period

2008–2013 they found that in most reviews, authors did not comment on applicability of results for primary care. It was concluded that, in order to support evidence synthesis relevant to primary care, review authors must do additional work to assess the applicability and relevance of study results to primary care practice. The authors suggest prospective registration of trials in primary care including specific information on applicability. In a commentary, [van Weel](#) emphasizes that in the translation of research to practice, the context in which patients are encountered should be addressed, not only considering health status but also socio-economic circumstances, cultural and religious features, social determinants of health, and the actual function in which the encounter between patient and professional takes place. The closer to the practice setting research data are collected, the more straightforward the translation is.

[Vucic and colleagues](#) analyzed whether protocols of Cochrane systematic reviews address data extraction from figures in included trials. After screening the protocols published between May 2013 and May 2014, the authors found that very few protocols mention data extraction from figures, and when mentioned the methods for extraction are unclear. They recommend to address this issue in the Cochrane Handbook and the methodological standards for reviews.

Less is not more when data in longitudinal studies are missing. This is an ongoing challenge researchers have to deal with, first and foremost to avoid missing data as much as possible, and in addition to minimize the impact of missing data on study results. [Jones c.s.](#) compared seven methods for the analysis of longitudinal studies with missing data due to drop-out or death, based on an longitudinal study of elderly women. While ignoring missing data resulted in biased estimates, it was shown that appropriate analytic methodology depends on the target of inference and the missing data mechanism. Therefore, for the primary analysis a method should be used that is robust to the most plausible missing data mechanism.

To adjust for confounding in observational research, propensity score (PS) and instrumental variable (IV) analyses are often used. These techniques are used also simultaneously in observational studies evaluating interventions. In a review of these studies [Laborde-Castérot and co-workers](#) analyzed the agreement between PS and IV results. They found that discrepancies are frequent. Researchers should carefully consider their analytic choices and readers must be cautious when interpreting results. The authors recommend further studies to clarify the roles of the two methods and to investigate the indications for each method.

A good balance of collecting sufficient but not too many data must also be found for clinimetric instruments. [Hutchings et al.](#) developed and validated the Gastrointestinal Symptom Rating Questionnaire (GSRQ) for patients with luminal gastrointestinal symptoms referred to secondary care

and in particular where no diagnosis has been made. The instrument was developed and piloted on patients with a known GI disorder, and tested in a sample of trial patients. The validation is reported to be successful in measuring health related quality of life (HRQL) in various GI conditions and where a formal diagnosis has not been established. The authors recommend the GSRQ to help monitor HRQL in patients before a diagnosis has been made and during the longitudinal course of their disease. [Gärtner and co-authors](#) validated the Labor and Delivery Index (LADY-X), a new delivery-specific utility measure reflecting the course of labor and birth, in a test-retest design among women who were surveyed online in the postpartum period. The LADY-X showed good reliability and construct validity and discriminated between groups. It may fulfill the need for a utility measure for cost-effectiveness studies for perinatal interventions. The comparability of the English, Chinese and Malay versions of the EQ-5D-5 L response labels, describing mild or extreme health problems, was investigated [Luo and colleagues](#) in primary care in Singapore. Interpretation and use of these labels varied among Singaporeans using different language versions of the instrument. They recommend further studies on ways to reduce the variations and to increase the cross-cultural measurement equivalence. Another cross-cultural adaptation was performed by [Wei and collaborators](#), who aimed to obtain an adapted and validated Chinese version of the Tampa Scale for Kinesiophobia (SC-TSK). Based on their study in LBP patients, the authors found good internal consistency, test-retest reliability, and construct validity. They conclude that the SC-TK can be considered a valid instrument for Chinese-speaking patients in mainland China. [Kopec and his group](#) developed and validated a multiattribute health (MAHU) utility scoring method for the CAT (Computer adaptive testing)-SD-QOL, measuring five domains of health-related quality of life. The empirical validation was carried out among retired persons. The study provided preliminary evidence to support the validity of the MAHU scoring system for the CAT-SD-QOL, which may be useful for economic evaluation studies. Future applications in clinical studies will provide further insight in the performance of the system.

The randomized registry trial, which is a trial built up on the structure of an existing registry of patients, can be useful to investigate volume-outcome relationships and learning curve effects, while better representing clinical reality, avoiding selection bias and promoting efficiency. This is highlighted by [Pieper and Neugebauer](#), who also discuss the preconditions registries have to meet for this purpose. [Papavasiliou and Payne](#) address another challenge when trying to observe reality: when studying the (evolution of) prevalence of continuous sleep sedation until death, as a type of end-of-life sedation, a problem is the gray area between sleep sedation until death and euthanasia. They discuss this issue in the context of the legal development in

various countries, potential sources of bias, and the need to promote the integrity of science, the art of medicine, and the betterment of public health.

J. André Knottnerus
Peter Tugwell
Editors

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

References

- [1] Weinstein M, Fineberg HV. Clinical decision analysis. Philadelphia: Saunders; 1980.
- [2] Haynes RB, Sackett DL, Tugwell P. Problems in the handling of clinical and research evidence by medical practitioners. *Arch Intern Med* 1983;143:1971–5.
- [3] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101–4.
- [4] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156–65.
- [5] Sawin VI, Robinson KA. Biased and inadequate citation of prior research in reports of cardiovascular trials is a continuing source of waste in research. *J Clin Epidemiol* 2015. pii: S0895–4356(15)00304-2.
- [6] Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;302:1109–17.
- [7] Nease RF Jr, Owens DK, Sox HC Jr. Threshold analysis using diagnostic tests with multiple results. *Med Decis Making* 1989;9:91–103.
- [8] Knottnerus JA, Buntinx F, van Weel C. General Introduction: evaluation of diagnostic procedures. In: Knottnerus JA, Buntinx, editors. *The evidence base of clinical diagnosis*. 2nd Ed. London: BMJ Books; 2008:1–19.
- [9] van Walraven C, Austin PC, Jennings A, Forster AJ. Correlation between serial tests made disease probability estimates erroneous. *J Clin Epidemiol* 2009;62:1301–5.
- [10] Oostenbrink R, de Wildt SN. Drug trials: Kids are no little adults and not all kids are the same. *J Clin Epidemiol* 2015. pii: S0895–4356(15)00333-9.