In the classic Hippocratic tradition, diagnosis, prognosis, and treatment are inherently connected around just one target: to improve the patient’s health [1]. And although scientific progress sometimes requires that researchers especially focus on one of the elements of this triad, they should always collaborate to integrate their work in guidelines and practices to serve that key target. This should make investigators alert to use any opportunity to anticipate this connection in their research. This sounds self-evident, but there are too many examples of highly sophisticated treatments being developed without appropriate protocol to pre-select the appropriate patient subgroups, and — probably even more — of advanced diagnostic technologies yielding detailed biomedical information without clear perspectives for better treatment [2]. As to the latter, the work by Siontis et al. is very important. In a systematic review of a large number of randomized controlled trials evaluating diagnostic interventions, they found that in only one in five there was evidence of significant changes in patient outcomes. Moreover, the effect of testing on patient outcomes was not correlated with further interventions, nor with diagnostic accuracy of tests. The authors recommend that diagnostic test evaluation should routinely include effects on various outcomes and that its reporting should be more patient-centered.

The clinical significance of randomized trials may also be less robust than reporting of statistically significant results suggest, unless the number of events is also taken into account, as is shown by Walsh and his group. In a review of RCTs reporting such results in high-impact medical journals, they analyzed how many events made the difference between a significant and non-significant result, using their newly developed ‘Fragility Index.’ As it was shown that in many RCTs significant results hinge on small numbers of events, the authors conclude that the Fragility Index complements the P-value and helps identify less robust results. The importance of paying attention to robustness of evidence is underlined by Alexander c.s., who reviewed a sample of the World Health Organization guidelines that had been rated according to the GRADE approach. The authors found that strong recommendations based on low or very low confidence estimates are very frequently made in WHO guidelines, and emphasize the need for further study on the reasons for such high uncertainty recommendations.

Considering appropriate presentation of evidence, inclusion and exclusion criteria should be transparent and well-justified. In a systematic review of trials on secondary prevention of cardiovascular events, Schmidt and his team studied the reporting and justification of exclusion criteria, and the extent to which exclusion criteria affected generalizability of study results. Virtually all trials reported the applied exclusion criteria, but only one of the 113 articles justified them. While there were no indications that generalizability was affected, the authors recommend better reporting of the reasons for using exclusion criteria. Also, inclusion of studies in systematic reviews is not always appropriately justified. For example, Ijaz and colleagues demonstrate in a review of all Cochrane reviews that considered nonrandomized studies that most did not justify including such studies. And when they do, most are not in line with Cochrane recommendations, while risk of bias assessment varies to a great extent. The authors provide recommendations for improvement.

Like criteria for inclusion and exclusion, blinding in randomized trials is often blindly used. Mathieu and coauthors revisited the justification of this principle in a theoretical analysis of the potential for bias in randomized trials even after successful blinding. They conclude that in order to fully eliminate bias, a state of complete ambivalence about allocation of every trial participant should be established. However, this may be difficult to achieve and may reduce generalizability of the trial’s findings.

Because problems with internal and external validity can never be excluded, in the discussion sections of research papers authors should always consider known and potential limitations of their work. In connection to work by Ioannidis [3], Yazici c.s., having the impression that in doing so basic science articles might perform less well than clinical research reports, compared the discussion sections of these two types of papers published in leading rheumatology journals. They found, indeed, a striking difference, and therefore make recommendations to promote self-critique, especially in basic science articles.

As to appropriate reporting, the abstracts of research papers also deserve attention. For the field of oncology this was evaluated by Ghimire and co-workers in a systematic review of phase III trials, using a score based on the CONSORT for Abstract guidelines and comparing the pre- and post-CONSORT periods. The investigators conclude that the reporting quality of RCT abstracts showed suboptimal improvement over time and that stricter adherence to the guidelines is needed.
Presentation of results of complex analysis can be supported by graphical tools to a great extent. Tan et al. show this for reporting network meta-analysis (NMA), with the aim of increasing accessibility, transparency, interpretability, and acceptability of such analyses. Using recommendations by agencies such as the National Institute for Health and Care Excellence, novel graphical approaches were designed. These tools can be tailored to display results relevant to the research question of interest, and targeted at a whole spectrum of users, from analysts to clinicians.

Patient-reported outcome measurement instruments have been recognized for decades now as key endpoints in clinical research [4]. In contrast to most other instruments, the Patient-Specific Functional Scale (PSFS) does not start from a preset generic or specific health-related content, but from preferences and priorities of individual patients. Based on data from a large series of consecutive physical therapy patients, Haxby Abbott et al. examined the validity of the PSFS, and concluded that it is an appropriate measure for statistical comparisons in clinical research. In evaluating patient reported outcome measurement instruments, insight in the minimal clinically important improvement (MCII) is essential. Ward and his team add to this insight, by examining baseline dependence of MCIIs for three measures for rheumatoid arthritis (RA) activity. They conclude that variation in MCII by baseline values is attributable to floor and ceiling effects rather than to expectations of particular patients. Minimally important change was also evaluated for the Manchester-Oxford Foot Questionnaire (MOXFQ). For this purpose, Dawson et al. used data from a prospective before-after study of patients undergoing foot or ankle surgery. Their findings can assist the interpretation of MOXFQ outcomes and can help researchers in this field to more precisely plan future studies.

Johnson and co-workers developed and evaluated the sensibility of a SSc (systemic sclerosis) specific instrument for use in a forced-choice study, conducted a forced-choice study to reduce and weight the criteria for SSc, and explored the agreement among experts on which patients are considered to have SSc. Following this approach, they defined a system that successfully reduced the number of candidate criteria, and produced a measure of the relative probability that a particular case has SSc. As the authors suggest, their work may serve as a template for development of classification criteria for other diseases.

In dealing with non-response in clinical epidemiological studies, safeguarding external validity by avoiding selective recruitment is one of the main objectives. In this context, David et al. evaluated substitution sampling as an alternative for more usual recruitment strategies such as sending reminders, using data from a prospective cohort study among diabetes patients. Their findings support substitution sampling as a recruitment option, as concerns about a higher risk of bias seem unwarranted. Santin and collaborators studied whether reweighting can correct for unit non-response in an occupational health surveillance survey, by linking survey data to administrative databases and sociodemographic data. The authors conclude that this approach can indeed effectively correct for nonresponse bias, while using sociodemographic data solely may not be sufficient for this purpose. As not only response at baseline but also continued retention in a study is of utmost importance, also from the perspective of efficient use of time and resources, the letter by Bennett Johns on and her group is very interesting. In a multinational study on possible environmental triggers of type I diabetes in children (the TEDDY Study), they developed a cumulative risk model to identify families most likely to leave the study within one year. Their approach may be used to focusing efforts to keep participants in a study, but also to define exclusion criteria at the time of study enrollment.

J. André Knottnerus

Peter Tugwell

E-mail address: anneke.germeraad@maastrichtuniversity.nl

(J.A. Knottnerus)

References