Journal of Clinical Epidemiology 112 (2019) v–vi

Journal of Clinical Epidemiology

EDITORIAL

Comprehensive evidence and clinical judgment

Avoiding 'type-of-data' bias

In empirical research, a lot of scientific information originates from questionnaires, measurement scales, lab tests and other technical investigations. That is, information that with little difficulty can be summarized in quantified expressions. However, in clinical consultations, a much broader variety of types of information and potentially relevant evidence is - often implicitly - considered. For example, experienced clinicians pay close attention to subtle non-verbal, physical or behavioral signs in patients, and changes in their social or environmental context. Such information can be important in the clinical assessment of individual patients [1].

More generally, clinical research should not be disproportionately focused only on types of information that can be easily harvested but must study the full spectrum of clinical information that may be relevant for making etiological interpretations, diaprognostic predictions, and individually targeting interventions. In other words, we should avoid what we might call 'type-of-data' bias, implying that important dimensions of real practice evidence are underused. By scientifically utilizing all types of potentially relevant information, we can bring the variety of evidence considered in practice closer to what we include in and make available from research. This enables us to test and validate the full spectrum of clinical information, and to provide a more comprehensive and representative evidence base for clinical judgment [2]. In addition, this helps improving guidelines and tools such as prediction rules to be more comprehensive, in support of better clinical decision making. A related challenge is evaluating clinical judgment in itself, which is essential in order to make this skill more transferable [3].

Clinical judgment and expert panels

Clinical judgment is often used to bridge gaps between available 'hard evidence' and the broader assessment needed for decision making. Van Houten et al. evaluated this approach in the field of diagnostic research: when a gold standard is lacking, expert panel diagnosis is often used as reference standard in diagnostic test accuracy studies. However, interobserver and intraobserver agreement is generally imperfect. These authors therefore evaluated the reproducibility of expert panel diagnosis for pediatric infectious diseases in febrile children, using the classification: bacterial infection, viral infection, or indeterminate. Diagnosis was reached when the majority of panel members came to the same conclusion. Intraobserver and intrapanel agreement was evaluated with 6 weeks and 3 years' time intervals, as the proportion of inconclusive diagnosis for a single, three-member, five-member, and seven-member expert panel. It was found that a panel consisting of three experts provided more reproducible diagnoses than an individual expert, while increasing panel size beyond three experts has no major advantage. The authors' general conclusion is that, as reference standard for infectious diseases during childhood, a panel diagnosis of experts is highly preferable to the diagnosis of a single expert. This work makes clear that, although not all ingredients of expert judgment are known, research can help to utilize it better, in the context of both clinical studies and practice.

Consensus and core outcome sets

Expert judgment is also important in developing core outcome sets (COS)and related consensus methods [4]. Gargon c.s. studied the standards of development for cancer core outcome sets, starting from the Core Outcome Set STAndards for Development (COS-STAD) which contains standards and criteria on scope, stakeholder involvement, and consensus process. With cancer being the disease area with the highest number of published COSs, the aim was to provide a baseline of cancer COS standards. Two reviewers independently assessed 49 identified COSs against the criteria. None of these met all criteria. While all studies met the four scope standards, only 16% met all three standards for stakeholders involved, and even less (4%) met all four standards for the consensus process. The investigators conclude that, with the exception of scope specification, there is much need for improvement. They add that poor reporting often made it difficult to assess whether standards were met. As the authors emphasize, their review may provide guidance on how to evaluate a published COS and can help users to assess whether a COS has been appropriately developed.

Reliability of risk of bias assessment

Expert judgment is required in the evaluation of risk of bias. As this process is subject to variation between experts, Minozzi and her team assessed the inter-rater reliability (IRR) and usability of the risk of bias in the nonrandomized studies of interventions tool (ROBINS-I) [5] now recommended for use in Cochrane Reviews. In a cross-sectional

https://doi.org/10.1016/j.jclinepi.2019.07.004
0895-4356

study, five raters independently applied ROBINS-I to nonrandomized cohort studies on vaccines, opiate abuse, and rehabilitation. In addition to calculating Fleiss' Kappa for multiple raters as a measure of IRR, the authors discussed difficulties and possible reasons for disagreement in applying the tool. In the 31 included studies, IRRs were 'slight' for overall judgment and individual domains. The main difficulties identified were the poor reporting of primary studies, misunderstanding of the questions in the tool, translation of questions into a final judgment, and incomplete guidance. The investigators recommend calibration exercises and intensive training before applying the tool in order to improve reliability. In addition, they emphasize that users must have good subject matter knowledge and much experience in the conduction of nonrandomized studies.

Saric et al. studied risk of bias assessments for selective reporting in Cochrane systematic reviews of randomized controlled trials. Risk of bias assessments were extracted, with judgments (low, high, or unclear risk) and supporting comments, and the sources of information mentioned in these comments were analyzed. The judgments were compared with the guidance from the Cochrane Handbook for Systematic Reviews of Interventions. It was found that at least 60% of the judgments for risk of selective reporting bias were not in line with the Cochrane Handbook. The authors therefore recommend interventions helping Cochrane authors to make adequate risk of bias assessments.

Improving knowledge dissemination

Not only in evidence production and assessment but also in evidence transfer comprehensive approaches should be used. In this context, the work of Sarkies et al. on video strategies to improve health professional knowledge is important. Using a novel 'helix- counterbalanced randomized controlled trial design' they compared the impact of delivering research information via video or written modalities with a no-information control, on professional knowledge of evidence provided in scientific journal articles. The interventions and data collection were organized via an online survey to nursing and allied health professionals in hospitals within a public health service. Alignment between respondent-perceived benefit of the intervention and article conclusions was the primary outcome. It was found that exposure to the video increased the likelihood of alignment compared to the no-information control, while this was not the case for exposure to the written modality. The authors conclude that use of video abstracts may be a useful adjunct to publishing research in dissemination activities. They recommend further research on the potential contribution of digital knowledge translation approaches toward more complex research implementation strategies and improving evidence uptake in health care policy and practice.

Methodological challenges

To conclude, in order to extend and strenghten the evidence base for health care practice all potentially relevant types of information must be used, and approaches to bridge gaps between available 'hard evidence' and what is required for decision making need attention. This is especially challenging where easily quantifiable data do not suffice and where clinical judgment becomes decisive. We would be interested to receive innovative clinical epidemiological work on methodological challenges in this field.

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

References

[1] Henry SG, Forman JH, Fetters MD. 'How do you know what Aunt Martha looks like?' A video elicitation study exploring tacit clues in doctor-patient interactions. J Eval Clin Pract 2011;17:933−9.
[2] Feinstein AR. Clinical judgment. Baltimore, MD: Williams and Wilkins Co.; 1967.
[3] Kassirer JP. Teaching clinical reasoning: case-based and coached. Acad Med 2010;85:1118−24.
[4] Tugwell P, Knottnerus JA. The need for consensus on consensus methods. J Clin Epidemiol 2018;99:vi−viii.
[5] Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.