Lost in translation between evidence and recommendations: expert opinion is needed to define "Level I"

Volovici, V.; Steyerberg, E.W.

Lost in Translation Between Evidence and Recommendations: Expert Opinion is Needed to Define "Level I"

***Victor Volovici*[1,2] *and Ewout W. Steyerberg*[2,3]**

INTRODUCTION

The aims of clinical practice guidelines (CPGs) are to provide clinicians with a comprehensive review of the best available evidence and to offer practical recommendations. Implementing the recommendations into clinical practice should reduce inappropriate variations and potential harms for patients, provided that the evidence is strong. During the early 2000s, attention was drawn to shortcomings in quality and reliability of CPGs. In 2011, the Institutes of Medicine (IOM) released a document aimed at clarifying the essential requirements necessary to build trustworthy CPGs.[1] As a result, the methodology behind guidelines development had come under increased scrutiny. The GRADE criteria, among others, aim to improve the creation of solid, reliable, nuanced CPGs.[2] In this Perspective article, we aim to reflect on the process of the creation of guidelines. We consider the consistency of recommendations in different guidelines (between-guideline consistency).We also consider within-guideline consistency (or durability[3]) as the amount of recommendations carried over from one edition to another in consecutive editions of the same CPG. We contrast 2 examples: guidelines for hypertension and guidelines for traumatic brain injury (TBI).

GUIDELINES FOR HYPERTENSION

Hypertension is a highly prevalent disease. Many randomized controlled trials (RCTs) have been performed, and the overall quality of studies is rather high, with 24 trials including 58,040 patients available for the choice of first-line therapy alone,[4] and 18 RCTs with a total of 141,807 participants for calcium-channel blockers.[5] A recent critical evaluation of hypertension CPGs across multiple specialty societies, however, revealed between-guideline inconsistency in the direction of recommendations on the same topics: some interventions were recommended by certain societies while discouraged by others.[6] The inconsistency in direction affected 35% of the high-importance recommendations, for example, recommendations for the choice of first-line therapy. No specific CPG nor a geographic factor was the key source of the inconsistency.

FACTORS CONTRIBUTING TO CPG BETWEEN-GUIDELINE INCONSISTENCY

One key factor is how the disease is defined. The American College of Cardiology/American Heart Association (ACC/AHA) guidelines lowered both the threshold used in the definition of the condition, as well as the target blood pressure for treatment. This change in definition has major consequences when applied at a nationwide level.[7]

The definition alone is not enough to account for the entire breadth of the between-guideline inconsistency, however: other possible explanations to the striking differences in recommendations include evidence selection and grading, the timing of the literature search, and stakeholder involvement.[6]

A guideline (CPG) developed and updated by a certain specialty society (such as the American Heart Association's Hypertension guideline), however, within-guideline consistency can be observed between consecutive editions. A study on the durability of ACC/AHA class I guidelines[3] has shown that 81% of these were carried over to a consecutive edition if they were based on high-quality evidence (>1 RCT).

GUIDELINES FOR TBI

Translating evidence into recommendations becomes considerably more difficult when diseases with a heterogeneous phenotype, a high degree of time-varying confounding and heterogeneous outcomes, such as severe TBI,[8] are the subject of CPGs. Despite massive research endeavors in severe TBI, including almost 200 randomized trials on various topics, the evidence base is weak.[9]

The between-guideline consistency in TBI is very low, under 5%. When we compare the TBI guidelines that are most often used, the Brain Trauma Foundation (BTF) Guidelines, with the American College of Surgeons Trauma Quality Improvement program TBI guidelines,[10] 75% of the level I and II recommendations exhibit between-guideline inconsistency (6 of the 8 total). Those that are consistent are based on the largest RCTs,[8,10]

Key words
- Clinical practice guidelines
- Evidence-based medicine
- Guideline consistency
- Hypertension
- TBI

Abbreviations and Acronyms
ACC/AHA: American College of Cardiology/American Heart Association
BTF: Brain Trauma Foundation
CPGs: Clinical practice guidelines
IOM: Institutes of Medicine
RCT: Randomized controlled trial
TBI: Traumatic brain injury

From the [1]*Department of Neurosurgery, Erasmus MC, Rotterdam;* [2]*Center for Medical Decision Making, Department of Public Health, Erasmus MC, Rotterdam; and* [3]*Department of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Center, Leiden, the Netherlands*

To whom correspondence should be addressed: Victor Volovici, M.D., Ph.D. [E-mail: v.volovici@erasmusmc.nl]

Citation: World Neurosurg. (2021) 150:39-41.
https://doi.org/10.1016/j.wneu.2021.03.095

which advocate to avoid the use of steroids and set some nutritional goals. Despite these recommendations being consistent between guidelines, the way the recommendations are formulated exhibits subtle nuances, which may lead to heterogeneous interpretations and implementation.

The within-guideline consistency in TBI is only 30%.[8] One reason for this low within-guideline consistency is that new available evidence overturned the results of older studies. The major factor, however, seems to be the increased methodological rigor with which studies were assessed.[8] Although laudable as such, this development hampered the translation of study results into clinically usable, relevant recommendations. The difference between TBI and hypertension in terms of within-guideline consistency is likely the result of the cardiology guidelines' heavier reliance on better, larger RCTs.

The TBI CPGs example highlights another aspect of guideline generation: as suggested in the IOM document,[1] a guidelines committee should include both methodologists, experts in interpreting evidence, as well as expert clinicians. Methodologists alone may interpret evidence properly but may not be able to contextualize the findings clinically, whereas clinicians alone may be subjected to biases, such as conflicts of interest or projection of own beliefs and experiences, to skew the interpretation of available evidence. In the case of TBI, one of the most commonly used interventions (hyperosmolar therapy) could not be the subject of any formal recommendation because of the lack of methodologically sound evidence.[8] The older recommendations were, however, restated in the text of the newest CPG (the BTF Guidelines). The committee judged the therapies to be very relevant clinically and added a warning that "they were formally no longer supported by the evidence."[8] This choice leaves clinicians without clear guidance and reflects the struggle between rigorous methodological assessment and perceived clinical importance of a topic. The leap between results of a study and issuing a recommendation is nonetheless essential to a proper "translation" of evidence. It may introduce between-guideline inconsistency because different groups interpret the same study results in a different manner.

EVIDENCE GENERATION AND GUIDELINE FORMULATION: CHALLENGES IN TRANSLATING SCIENCE TO PRACTICE

Obviously, lower-level recommendations (based mainly on observational studies) are more susceptible to interpretation than those supported by multiple RCTs, as illustrated in the earlier-mentioned examples. Still, between-guideline inconsistency is noted even for some recommendations issued on strong evidence.[6] This observation suggests that the process of CPG generation needs to improve.

Evidence generation and, as such, the whole concept of evidence-based practice, is becoming more complex. Sometimes, RCTs are criticized for sacrificing some external validity for strong internal validity.[11] One way forward would be to apply a "risk-modeling" approach, wherein a multivariable model predicts the risk for an outcome and is applied to disaggregate patients within RCTs to define risk-based variation in benefit.[12] A step further would be to create risk-based recommendations, bearing in mind that this approach would bring methodological challenges of its own.

At the present time, the translation of evidence into guidelines is not a purely scientific exercise, as illustrated by the hypertension and TBI CPGs. Practitioners find themselves in a situation in which following certain recommendations implies not following others, which also claim to be based on high levels of evidence. Guidelines are used by stakeholders and sometimes constitute grounds of claims in medical malpractice lawsuits. The heterogeneous interpretation of evidence when independent agencies evaluate the same evidence means that guidelines are failing in the same way that studies do when their findings are irreproducible. The GRADE criteria support issuing level I recommendations fitted to particular socioeconomic contexts. However, this argument is insufficient to explain why separate guideline committees in high-income countries judge a solid evidence base, encompassing multiple robust trials, and reach considerably different conclusions.

CONCLUSIONS

CPGs are essential tools for clinicians, stakeholders, researchers, policy makers, and, most importantly, carry direct, lasting impact on patients. Just as research, CPGs also need to have between-guideline and within-guideline consistency (akin to reproducibility of studies). Clinicians and researchers should consider the lower consistency of guidelines that are not based on at least one strong RCT. If not done carefully, guideline generation and "evidence translation" lead to recommendations that might be level I, grade A, de jure, but largely expert preference and opinion de facto.

REFERENCES

1. Institute of Medicine (US). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Graham R, Mancher M, Miller Wolman D, et al., editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209539/. Accessed January 1, 2021.
2. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13.
3. Neuman MD, Goldstein JN, Cirullo MA, Schwartz JS. Durability of class I American College of Cardiology/American Heart Association clinical practice guideline recommendations. *JAMA*. 2014;311:2092-2100.
4. Wright JM, Musini VM, Gill R. First-line drugs for hypertension. *Cochrane Database Syst Rev*. 2018;4:CD001841.
5. Chen N, Zhou M, Yang M, et al. Calcium channel blockers versus other classes of drugs for hypertension. *Cochrane Database Syst Rev*. 2010;8:CD003654.
6. Alper BS, Price A, van Zuuren EJ, et al. Consistency of recommendations for evaluation and management of hypertension. *JAMA Netw Open*. 2019;2:e1915975.
7. Ioannidis JPA. Diagnosis and treatment of hypertension in the 2017 ACC/AHA guidelines and in the real world. *JAMA*. 2018;319:115-116.
8. Volovici V, Steyerberg EW, Cnossen MC, et al. Evolution of evidence and guideline recommendations for the medical management of severe traumatic brain injury. *J Neurotrauma*. 2019;36:3183-3189.

9. Bragge P, Synnot A, Maas AI, et al. A state-of-the-science overview of randomized controlled trials evaluating acute management of moderate-to-severe traumatic brain injury. *J Neurotrauma*. 2016;33:1461-1478.

10. American College of Surgeons. ACS TQIP best practice in the management of traumatic brain injury. 2019. Available at: https://www.facs.org/quality-programs/trauma/tqp/center-programs/tqip/best-practice. Accessed April 14, 2020.

11. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-1892.

12. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Ann Intern Med*. 2020;172:35-45.

Conflict of interest statement: The authors declare that the article content was composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Citation: World Neurosurg. (2021) 150:39-41.
https://doi.org/10.1016/j.wneu.2021.03.095

Journal homepage: ***www.journals.elsevier.com/world-neurosurgery***

Available online: ***www.sciencedirect.com***