RESEARCH

Assessing the process and outcome of the development of practice guidelines and recommendations: PANELVIEW instrument development

Wojtek Wiercioch MSc, Elie A. Akl MD PhD, Nancy Santesso MLIS PhD, Yuan Zhang PhD, Rebecca L. Morgan MPH PhD, Juan José Yepes-Nuñez MD PhD, Sérgio Kowalski MD PhD, Tejan Baldeh MPH, Reem A. Mustafa MD PhD, Kaja-Triin Laisaar MD PhD, Ulla Raid MSc, Itziar Etxeandia-Ikobaltzeta PharmD, Alonso Carrasco-Labra DDS PhD, Matthew Ventresca MSc MBA, Ignacio Neumann MD PhD, Maicon Falavigna MD PhD, Romina Brignardello-Petersen DDM PhD, Gian Paolo Morgano PHd, Jan Brożek MD PhD, Meghan McConnell PhD, Holger J. Schünemann MD PhD; for the PANELVIEW Working Group*

Cite as: *CMAJ* 2020 October 5;192:E1138-45. doi: 10.1503/cmaj.200193

*A list of the members is provided at the end of the article.

ABSTRACT

BACKGROUND: Guideline recommendations may be affected by flaws in the process, inappropriate panel member selection or conduct, conflicts of interest and other factors. To our knowledge, no validated tool exists to evaluate guideline development from the perspective of those directly involved in the process. Our objective was to develop and validate a universal tool, the PANELVIEW instrument, to assess guideline processes, methods and outcomes from the perspective of the participating guideline panellists and group members.

METHODS: We performed a systematic literature search and surveys of guideline groups (identified through contacting international organizations and convenience sampling of working panels) to inform item generation. Subsequent groups of guideline methodologists and panellists reviewed items for face validity and missing items. We used surveys, interviews and expert review for item reduction and phrasing. For reliability assessment and feedback, we tested the PANELVIEW tool in 8 international guideline groups.

RESULTS: We surveyed 62 members from 13 guideline panels, contacted 19 organizations and reviewed 20 source documents to generate items. Fifty-three additional key informants provided feedback about phrasing of the items and response options. We reduced the number of items from 95 to 34 across domains that included administration, training, conflict of interest, group dynamics, chairing, evidence synthesis, formulating recommendations and publication. The tool takes about 10 minutes to complete and showed acceptable measurement properties.

INTERPRETATION: The PANELVIEW instrument fills a gap by enabling guideline organizations to involve clinicians, patients and other participants in evaluating their guideline processes. The tool can inform quality improvement of existing or new guideline programs, focusing on insight into and transparency of the guideline development process, methods and outcomes.

As a product of a group process that involves project planning, synthesizing evidence and deliberation by guideline group members to reach consensus and formulate recommendations, health guidelines are highly influential in determining practice.[1,2] Guideline development also requires careful coordination of multiple teams with specialized knowledge.[1,3,4] These teams typically include an oversight committee responsible for project planning, working groups responsible for preparation and technical aspects of evidence synthesis, and a guideline panel tasked with prioritizing questions and formulating recommendations.

Guideline group processes may be prone to influence by people with strong opinions, imbalanced group member

© 2020 Joule Inc. or its licensors

characteristics or unqualified members, or may not use the best available evidence.[2,5–8] Currently available instruments assessing the trustworthiness of practice guidelines rely on what the guideline authors report, typically in peer-reviewed publications, or reports of organizations or manuals.[9,10] However, what authors report may be generic or incomplete, lack transparency or be inconsistent with the assessment of all group members, and what is reported may not always reflect what happened.[11] For example, the Appraisal of Guidelines, Research and Evaluation (AGREE)[9] and Reporting Items for Practice Guidelines in Healthcare (RIGHT)[10] tools appropriately call for conduct of systematic reviews as part of guideline development and appropriate disclosure of potential influence of conflicts. Although systematic reviews should inform guidelines, their conduct guarantees neither their quality nor that they are used appropriately by the panel for making recommendations. Likewise, having a conflict-of-interest declaration and management policy for guidelines does not necessarily guarantee that conflicts are well managed when guideline groups make recommendations.

Existing tools do not evaluate essential steps and processes, such as giving appropriate consideration to the evidence and ensuring that all panel members have an equal voice, as they take place.[9,10,12] An internal evaluation by participating guideline group members would provide this valuable insight. In addition, ensuring that panel members view the process as appropriate and one that results in a credible guideline will help ensure they see value in their contribution. By obtaining an assessment from the participants, guideline developers could identify areas of their processes participants view as needing improvement, as well as dissenting views among participants. They could then use this information to modify their methods and approaches, and to ensure the credibility of their guidelines and the trustworthiness of the recommendations.

The objective of this research was to develop and validate a tool for assessing guideline panel members' perception of the appropriateness of, and satisfaction with, the process, methods and outcome of the development of a health guideline.

Methods

For the development of the instrument, which we named PANELVIEW, we followed methods for scale development, including item generation based on existing literature, item reduction through key informant and expert feedback and consensus and field testing with guideline panels (Table 1).[13] At each step, participating panel members and guideline methodologists drawn from organizations from diverse geographic areas who produced guidelines about different clinical topics, assessed information about the items to evaluate the appropriateness of guideline development.

Table 1: Overview of steps and participants in the PANELVIEW tool development

Step	Method	Sources and participants	No. of resulting items
1. Item generation	Systematic review	17 articles	694
	Contacting guideline developers	3 source documents from 19 developers (AAO-HNS, ACP, CAR, CCO, CDC, Chile MoH, Colombia MoH, CTFPHC, DCGP, FMSD, KP, NHMRC, NICE, NKCHS, OPHA, RKIG, SIGN, South Africa MoH, USPSTF)	
	Key informant survey	62 panellists: 13 panels (WHO, KSA MoH)	
2. Item reduction	Key informant survey	9 panellists and guideline methodologists: 3 panels (WHO, WAO, CCO)	95
	Key informant interview	13 panellists and guideline methodologists: 2 panels (WAO, EHIF)	
	Expert review and consensus	Study team experts and guideline methodologists	
3. Response phrasing	Key informant survey	26 panellists and guideline methodologists: 3 workshops, 1 panel (AGA)	34
4. Field testing	Use of tool with working panels	12 panellists, 1 pilot panel (NHF) 94 panellists, 8 panels (AABB,* ADA, RA Adaptation, RARE-Bestpractices, WHO)	

Note: AAO-HNS = American Academy of Otolaryngology–Head and Neck Surgery, ACP = American College of Physicians, ADA = American Dental Association, AGA = American Gastroenterological Association, CAR = Canadian Association of Radiologists, CCO = Cancer Care Ontario, CDC = Centers for Disease Control and Prevention, CTFPHC = Canadian Task Force on Preventive Health Care, DCGP = Dutch College of General Practitioners, EHIF = Estonian Health Insurance Fund, FMSD = Finnish Medical Society Duodecim, KP = Kaiser Permanente, KSA = Kingdom of Saudi Arabia, MoH = Ministry of Health, NHF = National Hemophilia Foundation, NHMRC = National Health and Medical Research Council Australia, NICE = National Institute for Health and Care Excellence, NKCHS = Norwegian Knowledge Centre for the Health Services, OPHA = Ontario Public Health Agency, RA Adaptation = Rheumatoid Arthritis Guideline Adaptation for the Eastern Mediterranean Region, RKIG = Robert Koch Institute Germany, SIGN = Scottish Intercollegiate Guidelines Network, USPSTF = US Preventive Services Task Force, WAO = World Allergy Organization, WHO = World Health Organization.
*Formerly the American Association of Blood Banks.

Item generation

Item generation began with discussion by 2 investigators (W.W., H.J.S.) of key domains for capturing the evaluation of guideline-related processes based on domains in the GIN-McMaster Guideline Development Checklist.[1] We hypothesized that all parts of the process might be relevant for assessing appropriateness and satisfaction of panel members.

We then conducted a systematic literature search to identify steps and themes in guideline development that relate to the appropriateness of the process. We searched MEDLINE and Embase from inception to November 2018 to identify studies that discussed or evaluated steps of guideline development.[1] We used controlled vocabulary and keywords to capture evaluation of the guideline development process and panel member perceptions (see Appendix 1, Supplemental Figure S1 and Supplemental Box S1, available at www.cmaj.ca/lookup/doi/10.1503/cmaj.200193/tab-related-content, for additional details).

To supplement our literature search, we contacted a convenience sample of 19 key informants, identified in a previous project,[1] who represented major guideline-development organizations globally. We asked whether the organizations currently conducted internal evaluation of their guideline-development processes or used specific tools (Table 1, step 1). For each step involving key informants, we used a new sample of participants as a method of confirming data and views obtained in the preceding step, and to ensure broad representation of views and perspectives.

Panel surveys

We surveyed members of 13 guideline panels to obtain primary data (Table 1, step 1). After their panel meetings had adjourned, 62 panellists completed hard-copy surveys consisting of 6 open-ended questions inquiring about the factors that affected their satisfaction and perception of the appropriateness of the process (Appendix 1, Supplemental Table S1, Supplemental Figure S2). We included the survey responses as a source document for data abstraction.

Data abstraction

We developed and pilot tested a structured data-abstraction form. Study team members (Y.Z., R.L.M., K.-T.L., U.R., M.V., J.J.Y.-N., R.A.M., N.S., S.K., T.B.) reviewed full texts of source documents and abstracted independently and in duplicate items that related to the appropriateness of the methods or processes of guideline development, panel members' views about methods or processes, or panel members' satisfaction. Supporting quotations from the source document were included for each item, along with proposed themes for grouping of items (e.g., conflict-of-interest management, training, group interaction).

A subgroup of the study team members (W.W., Y.Z., R.L.M., S.K., J.J.Y.-N.) independently identified and merged duplicate items (i.e., items that measured or asked about the same aspect of the guideline-development process). The decisions were assessed by a second reviewer and discussed in a team meeting during which we finalized the deduplication, initial item phrasing and allocation to specific themes of the guideline-development process.

Item reduction

Feedback from key informants

We surveyed and conducted interviews with a convenience sample of 22 key informants, including guideline developers, methodologists and panel members, to obtain feedback about the initial list of items (Table 1, step 2; Appendix 1, Supplemental Table S2). Respondents were asked to rate on a 7-point Likert-type scale ranging from 1 (not important) to 7 (very important) how important they considered each of the items to be for evaluating the guideline-development process, to suggest modifications and to identify any missing items. We also sought to obtain in-depth feedback about the initial list and asked participants to comment on the level of detail, clarity and redundancy in the items. We then sent the list of items to participants for review in advance and conducted interviews in person at guideline panel meetings in presence of a note taker. We pilot tested both the survey and the interview guide.

Study team review and consensus meeting

Concurrently with the key informant surveys and interviews, we provided study team members (N.S., I.E.I., Y.Z., K.-T.L., S.K., R.B.-P., M.V., M.F., G.P.M.) with a structured feedback form to review the initial list of items and provide suggestions for modifications and theme categorization, or to suggest potentially missing items. For each item, we summarized the study team's suggestions, and the key informants' feedback and rating of importance. Study team members reviewed the summary and individually suggested to keep, modify, merge or delete items in preparation for a consensus meeting, during which the study team finalized decisions about each item based on discussion and group consensus. We then refined item wording based on our experience with developing other measurement instruments.[14,15]

Phrasing of items and response options

We conducted surveys with 26 additional key informants to determine the phrasing of items and response options (Table 1, step 3; Appendix 1, Supplemental Table S2). We asked respondents to indicate their preference for one of three 7-point Likert-type options regarding phrasing of responses and items, presenting 8 example items from the tool. The first option asked about appropriateness, the second about satisfaction, and the third, representing the original Likert scale, about agreement with the topic presented by the item.

Testing with panels

After item reduction, we pilot tested the PANELVIEW tool with 1 guideline panel consisting of 12 members. Subsequently, we made minor revisions to clarify item wording and the order of items in the tool and then used the tool with an additional 8 guideline panels consisting of 94 panellists (Table 1, step 4; Appendix 1, Supplemental Table S3). Panel members completed the PANELVIEW survey individually, expressing their agreement with each survey item on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) (e.g., "There was appropriate management of potential bias in panel members' interpretation of evidence and

alignment with prior beliefs"). Panel members also provided feedback on the clarity of the instructions, clarity of items and the survey length.

Analysis

We used generalizability theory (G theory), a method for evaluating the reliability of measurements, to assess the reliability of scores obtained across the different panel groups.[13] We calculated the item mean scores, standard deviations (SDs) and ranges across individual panellists. For individual panellists, we calculated overall scores as the mean of their item ratings. We conducted the preliminary reliability analyses at the individual panellist level as well as at the panel level, whereby we obtained item means by collapsing across individual panellists. We estimated multiple sources of variance (*G*), including the respondent, panel, item and domain, using a nested G theory study.[13] The guideline-development process of different panel groups served as the object of measurement, individual respondents were nested within panels, and individual items were nested within the PANELVIEW survey domains (Appendix 1, Supplemental Table S5, Supplemental Figure S5).

Ethics approval

The Hamilton Integrated Research Ethics Board approved this study before data collection.

Results

Our systematic literature search, contact with key informants and surveys of 13 guideline panels yielded 17 published articles,[6,16–31] 3 additional source documents (Appendix 1, Supplemental Figure S3) and 62 survey responses. We abstracted a list of 694 items, which, after evaluation and deduplication, resulted in 95 items grouped across 17 themes covering guideline development (Appendix 1, Supplemental Table S4).

Informed by the rating of importance of the 95 items and feedback from key informants, we removed 23 items that scored low on importance as part of our consensus process. We merged 38 items with other items considered redundant. We phrased each item to ensure it assessed only 1 component of the guideline-development process. The final list included 34 items. Of the 26 key informants surveyed about the phrasing of response options, 19 (73%) indicated preference for using the Likert scale.

Generalizability and reliability

The analysis of variance from the nested G study showed an overall test reliability coefficient of 0.35 (Appendix 1, Supplemental Table S5). This result is likely an effect of enrolling homogeneous panels with regard to processes and methods. The tool's domains and individual items within the domains each accounted for 4% of the variance, which also suggests that the processes for the guideline efforts we evaluated were similar across the domains and items. The guideline panels accounted for 28% of the variance, and participants within panels accounted for 55% of the variance in scores. This indicates that variation was captured in panellists' assessments between the guideline panels and in panellists' ratings within the panels. Despite the homogeneity of the groups, the tool was able to identify varying views of guideline panel members indicating higher and lower satisfaction or perception of appropriateness.

Response variation, item–item correlation and internal consistency

Within the panels, mean scores for items ranged from 4.0 to 7.0, item–item correlation values ranged from –0.76 to 0.96, and item–total correlation values ranged from –0.17 to 0.89. Across the 8 panels, the mean scores for items ranged from 5.5 to 6.8 (Table 2). There was high internal consistency in rating of satisfaction and appropriateness of the process within the 8 panels, with Cronbach's α ranging from 0.85 to 0.98 (Table 3). For individual panellists, item responses ranged from 1 to 7, and item–item correlation values ranged from 0.003 to 0.719. This suggests, on an individual respondent level, that the tool distinguished between responses and that there was no end-of-scale aversion. Item–total correlation values by individual raters ranged from 0.40 to 0.80, which suggests that the items were measuring different aspects of the guideline process.

Feedback from guideline panel group members

Respondents reported that they did not have difficulty completing the questionnaire (mean rating on Likert scale 6.4 [SD 0.6]). Respondents, on average, felt that the questionnaire was neither too long nor too short (mean rating 3.5 [SD 1.7], with a rating < 4 suggesting that the questionnaire was not too long). For 68 respondents who completed the survey online, the average completion time was 12 (SD 7) minutes, and the median time was 10 minutes (we excluded 12 respondents with a recorded completion time of ≥ 30 min, who presumably took a break while completing the questionnaire). In response to a suggestion from 8 respondents, we added an option to respond to relevant items as "not applicable" (e.g., to allow panel chairs to skip items that request evaluation of their chairing of the panel). The final PANELVIEW tool is available at https://heigrade.mcmaster.ca/guideline-development/panelview.

Interpretation

We developed a tool, the PANELVIEW instrument, that allows guideline developers to assess their processes, methods and outcomes by directly involving clinicians, patients and any other guideline group member in the evaluation.

Existing instruments for assessing guideline credibility rely on the guideline authors' report, which may describe the process as planned but not as implemented or as viewed by all group members, and may not reflect all relevant nuances of the process that affect the trustworthiness of recommendations. The PANELVIEW tool focuses on these important nuances and on the transparency of the guideline-development process, allowing organizations responsible for guideline development to inform their quality-improvement efforts.

Table 2 (part 1 of 2): Mean scores across panels for the PANELVIEW tool

Domain	Item	Mean score ± SD*	Range*
1. Administration	1 - The logistical support provided for organization of the guideline project and panel meeting was appropriate (e.g., scheduling of meeting, sharing of materials, venue/location)	6.29 ± 0.86	3–7
	2 - There was adequate preparatory work and meetings/teleconferences before the final panel meeting	5.79 ± 1.34	2–7
	3 - Adequate time was given for guideline group members to complete tasks (e.g., surveys, providing feedback) throughout the development of the guideline and to review the evidence summary and other material before the panel meeting	5.87 ± 1.12	2–7
	4 - Adequate time was allotted for the final panel meeting for all guideline questions to be discussed and recommendations to be formulated	5.54 ± 1.49	2–7
	5 - The panel meeting had a clearly defined agenda and objectives	6.48 ± 0.87	3–7
2. Training	6 - Information was provided about the specific methodology and frameworks to ensure understanding of the overall process and steps that would be used to develop the guideline	6.35 ± 0.81	3–7
3. Panel chair	7 - The panel chair(s) was able to provide clinical and methodological guidance during the meeting, providing direction and support for decision-making	6.56 ± 0.78	3–7
	8 - The panel chair(s) was able to manage the group process, establishing an atmosphere of support that ensured involvement of all panel members in the discussion and free expression of opinions	6.60 ± 0.63	5–7
4. Conflict of interest	9 - There was appropriate management of potential interests (financial, academic) of guideline group members, of the organization and in the evidence synthesis being free from bias	6.23 ± 1.11	1–7
	10 - There was appropriate management of potential bias in panel members' interpretation of evidence and alignment with prior beliefs	5.97 ± 1.20	1–7
5. Scoping the guideline	11 - The panel was given sufficient opportunity to be involved in the prioritization of questions and scoping of the guideline	6.26 ± 0.84	4 -7
	12 - The final scope of the guideline was clearly communicated to the guideline group and agreement was sought	6.32 ± 0.83	4–7
6. Methodology and process	13 - The evidence synthesis was rigorous	6.14 ± 0.97	3–7
	14 - A transparent and usable summary of the evidence was made available for the discussion	6.23 ± 1.14	2–7
7. Considering the evidence and contributing through expertise	15 - Appropriate consideration was given to the evidence, including all relevant types, and balanced with panel members' input and opportunity to use their experience to interpret the evidence	6.45 ± 0.71	3–7
	16 - The method or process used for decision-making with the available evidence was appropriate	6.40 ± 0.74	3–7
	17 - There was appropriate involvement and consultation with key stakeholders during the guideline development	5.84 ± 0.97	4–7
	18 - Appropriate consideration was given to patients' views, perspectives, values and preferences	5.71 ± 1.20	2–7
8. Formulating the recommendations	19 - An appropriate method was used for formulating the recommendations with transparency of judgments made	6.45 ± 0.77	3–7
	20 - Appropriate consideration was given to relevant external factors (e.g., policy implications, setting-specific health care factors, acceptability of recommendations) in formulating the guideline recommendations	6.12 ± 0.88	2–7
	21 - The consensus method used by the panel was appropriate, allowing ability to reach consensus	6.36 ± 0.72	4–7
	22 - The wording of the guideline recommendations formulated was clear and actionable	6.26 ± 0.79	4–7
	23 - There was transparency in going from the panel's recommendations to the final recommendations that appear in the guideline report, and notice was given about any changes made	6.24 ± 0.99	4–7

Table 2 (part 2 of 2): PANELVIEW tool mean scores across panels

Domain	Item	Mean score ± SD*	Range*
9. Group composition	24 - There was diversity in membership and adequate representation of backgrounds, specialties and balance of expertise in the panel composition	6.35 ± 0.88	3–7
	25 - The panel size was appropriate	6.41 ± 0.81	3–7
10. Group roles	26 - The required commitment was at an appropriate level for the guideline group members	6.47 ± 0.67	4–7
	27 - The contributions of the guideline group members were valued and appropriate credit was given	6.52 ± 0.68	4–7
11. Group interaction	28 - There was mutual respect between guideline group members, with friendly and professional conduct	6.71 ± 0.54	5–7
12. Implementation and dissemination planning	29 - Appropriate consideration was given to the discussion of research gaps and needs for future research	6.28 ± 0.85	3–7
	30 - Appropriate consideration was given for the planning of dissemination and implementation of the guideline	6.08 ± 1.04	3–7
13. Writing guideline	31 - The writing of the guideline was well planned, with agreement on the format(s) and opportunity for panel members to provide input and review the guideline draft	5.90 ± 1.19	2–7
14. Incentive	32 - I felt that my involvement in the guideline will have an impact on the health of people	6.30 ± 0.81	4–7
15. Overall satisfaction	33 - Overall, I was satisfied with the guideline development process	6.48 ± 0.69	4–7
	34 - I would participate in this guideline development process again	6.78 ± 0.44	5–7

Note: SD = standard deviation.
*On a Likert scale ranging from 1 to 7.

We followed best practice for instrument development, including reviewing the literature, contacting key informants at guideline organizations and surveying panellists about key factors affecting guideline development. We tested the tool successfully with panels from international guideline organizations.

The PANELVIEW instrument is designed to identify strengths and weaknesses of a guideline-development group's process and methods in a structured manner, and highlight specific areas for improvement as identified by the participants by assessing ratings within individual domains. The tool enables evaluation of guideline development by participating group members in its entirety or in phases. How the guideline process is organized may differ between organizations, for example, between those that convene 1 final panel meeting and those that maintain a standing panel with repeated meetings. This will determine whether developers administer the PANELVIEW tool once at the conclusion of a guideline project or throughout the process as the steps take place.

Table 3: PANELVIEW tool mean scores and internal consistency across guideline panels

Guideline panel	Mean score ± SD*	Cronbach α
1	6.46 ± 0.32	0.92
2	6.04 ± 0.43	0.98
3	6.53 ± 0.28	0.88
4	6.05 ± 0.59	0.96
5	6.27 ± 0.53	0.95
6	6.07 ± 0.34	0.96
7	6.37 ± 0.24	0.95
8	6.01 ± 0.50	0.85

Note: SD = standard deviation.
*On a Likert scale ranging from 1 to 7.

The PANELVIEW instrument is not intended to replace existing tools that offer guidance on the appropriate steps for guideline development or assess the credibility of published guideline reports. It offers an approach for identification of issues in the guideline-development process and methods by those who participate in it or directly observe it, such as technical experts and methodologists. The tool can serve to inform evaluation or quality improvement of new or existing guideline programs, respectively.

The rigour of development with the end-user in mind is the main strength of our work. First, we applied item-generation methods drawing on multiple sources: literature, contacting organizations, panel surveys and a team with extensive experience in the guideline field. Second, we involved other key informants from multiple organizations and participation on panels for input on items and face validity, allowing data saturation. Third, we field tested the tool with groups that focused on a variety of guideline topics.

We plan to administer the PANELVIEW tool with additional, diverse panels from various guideline organizations. Guideline organizations can access the tool at https://heigrade.mcmaster.ca/guideline-development/panelview to participate. We will seek further feedback on use of the tool, for example, about the potential for public reporting of PANELVIEW assessments to increase transparency. The high Cronbach α coefficients may indicate the presence of redundant items. Sampling of more panels will allow us to assess whether any refinement of tool items is necessary and conduct factor analysis for further evaluation of tool domains. Additional opportunities include comparative studies, for example, comparing PANELVIEW assessments of panellists to those of other group members (e.g., nonvoting observers), as well as evaluating global ratings and judgments of panel success and guideline credibility against ratings of the tool.

Limitations

A potential limitation of our research is that we did not conduct systematic searches of the nonmedical literature in the areas of business, education and policy-making for relevant items. At each step involving key informants, we used convenience sampling, which may introduce sampling bias. To address this, we drew on a broad representation of working guideline panellists, with varying levels of experience, as well as guideline-development experts from organizations representing a wide range of processes and methods.

The 8 guideline groups involved in field testing the PANELVIEW tool were recruited through key informants, and, for some aspects of development, the groups used similar methods (e.g., using the GRADE approach for assessing quality of evidence and strength of recommendations) and involved experienced group chairs. The high scores on many items and the lower overall reliability coefficient of 0.35 to discriminate between groups indicated that the groups were likely all high performing. Despite this, we observed variability in scores within the groups, which would allow guideline developers to identify whether individual panellists viewed the process and specific aspects of the process as more or less appropriate.

Conclusion

The PANELVIEW instrument allows capturing of panellists' perspectives when they participate in guideline development, which addresses a gap in the field. Given the importance of guidelines and their impact on recipients and providers of care, optimizing the quality of their development is a logical step.

References

1. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 2014;186:E123-42.
2. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 5. Group processes. *Health Res Policy Syst* 2006;4:17.
3. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: Graham R, Mancher M, Miller Wolman D, et al., editors. *Clinical practice guidelines we can trust*. Washington: National Academies Press; 2011.
4. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;156:525-31.
5. Schünemann HJ, Woodhead M, Anzueto A, et al. A guide to guidelines for professional societies and other developers of recommendations: introduction to integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proc Am Thorac Soc* 2012;9:215-8.
6. Neumann I, Karl R, Rajpal A, et al. Experiences with a novel policy for managing conflicts of interest of guideline developers: a descriptive qualitative study. *Chest* 2013;144:398-404.
7. Boyd EA, Bero L. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. *Health Res Policy Syst* 2006;4:16.
8. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Res Policy Syst* 2006;4:15.
9. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839-42.
10. Chen Y, Yang K, Marušić A, et al. A reporting tool for practice guidelines in health care: the RIGHT statement. *Ann Intern Med* 2017;166:128-32.
11. Alexander PE, Li SA, Gionfriddo MR, et al. Senior GRADE methodologists encounter challenges as part of WHO guideline development panels: an inductive content analysis. *J Clin Epidemiol* 2016;70:123-8.
12. Shiffman RN, Shekelle P, Overhage JM, et al. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med* 2003;139:493-8.
13. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 5th ed. Oxford (UK): Oxford University Press; 2015.
14. Puhan MA, Behnke M, Laschke M, et al. Self-administration and standardisation of the chronic respiratory questionnaire: a randomised trial in three German-speaking countries. *Respir Med* 2004;98:342-50.
15. Schünemann HJ, Goldstein R, Mador MJ, et al. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. *Eur Respir J* 2005;25:31-40.
16. Atkins L, Smith JA, Kelly MP, et al. The process of developing evidence-based guidance in medicine and public health: a qualitative study of views from the inside. *Implement Sci* 2013;8:101.
17. Calderón C, Rotaeche R, Etxebarria A, et al. Gaining insight into the clinical practice guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. *BMC Health Serv Res* 2006;6:138.
18. Chang LW, Kennedy CE, Kennedy GE, et al. Developing WHO guidelines with pragmatic, structured, evidence-based processes: a case study. *Glob Public Health* 2010;5:395-412.
19. Gardner B, Davidson R, McAteer J, et al. A method for studying decision-making by guideline development groups. *Implement Sci* 2009;4:48.
20. Harding E, Brown D, Hayward M, et al. Service user perceptions of involvement in developing NICE mental health guidelines: a grounded theory study. *J Ment Health* 2010;19:249-57.
21. Hopthrow T, Feder G, Michie S. The role of group decision making processes in the creation of clinical guidelines. *Int Rev Psychiatry* 2011;23:358-64.
22. Hutchings A, Raine R. A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. *J Health Serv Res Policy* 2006;11:172-9.
23. Hutchings A, Raine R, Sanderson C, et al. An experimental study of determinants of the extent of disagreement within clinical guideline development groups. *Qual Saf Health Care* 2005;14:240-5.
24. Khodyakov D, Hempel S, Rubenstein L, et al. Conducting online expert panels: a feasibility and experimental replicability study. *BMC Med Res Methodol* 2011;11:174.
25. Moreira T, May C, Mason J, et al. A new method of analysis enabled a better understanding of clinical practice guideline development processes. *J Clin Epidemiol* 2006;59:1199-206.
26. O'Brien KK, Wilkins A, Zack E, et al. Developing clinical practice guidelines in HIV rehabilitation: process recommendations and guiding principles. *AIDS Educ Prev* 2011;23:457-68.
27. Pagliari C, Grimshaw J. Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. *J Eval Clin Pract* 2002;8:145-53.
28. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. *J Eval Clin Pract* 2001;7:165-73.
29. Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. *BMJ* 2005;331:631-3.
30. Raine R, Sanderson C, Hutchings A, et al. An experimental study of determinants of group judgments in clinical guideline development. *Lancet* 2004;364:429-37.
31. Williams W, Lowery NE, Lyalin D, et al. Development and utilization of best practice operational guidelines for immunization information systems. *J Public Health Manag Pract* 2011;17:449-56.

Competing interests: Authors of this manuscript have been involved in the development of various guideline manuals that are referenced in this article. Wojtek Wiercioch, Elie Akl and Holger Schünemann will retain copyright of the PANELVIEW tool through McMaster University (Technology ID 20-031), and the tool may be licensed in the future. Maicon Falavigna received consulting fees from Novartis, AbbVie, Roche, Boehringer Ingelheim, PTC Therapeutics and Sanofi Genzyme, unrelated to the submitted work. No other competing interests were declared.

This article has been peer reviewed.

Affiliations: Department of Health Research Methods, Evidence, and Impact (Wiercioch, Akl, Santesso, Zhang, Morgan, Baldeh, Mustafa, Etxeandia-Ikobaltzeta, Carrasco-Labra, Ventresca, Brignardello-Petersen, Morgano, Brożek, Schünemann), McMaster University, Hamilton, Ont.; Department of Internal Medicine (Akl), American University of Beirut, Beirut, Lebanon; School of Medicine (Yepes-Nuñez), Universidad de los Andes, Bogotá, Colombia; Department of Internal Medicine (Kowalski), Federal University of Paraná, Curitiba, Brazil; Division of Nephrology and Hypertension (Mustafa), Department of Medicine, University of Kansas Medical Center, Kansas City, Kan.; Guideline Development Group (Laisaar), Faculty of Medicine, Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia; Ministry of Social Affairs (Raid), Tallinn, Estonia; Department of Internal Medicine (Neumann), Pontificia Universidad Católica de Chile, Santiago, Chile; Federal University of Rio Grande do Sul (Falavigna), Institute for Health Technology Assessment, Porto Alegre, Brazil; Department of Medicine (Brożek, Schünemann), McMaster University, Hamilton, Ont.; Department of Innovation in Medical Education (McConnell), University of Ottawa, Ottawa, Ont.

PANELVIEW Working Group members: Yngve Falck-Ytter MD, Division of Gastroenterology, Case Western Reserve University Cleveland, Ohio.; Iván D. Florez MD PhD, Department of Pediatrics, Universidad de Antioquia, Medellin, Colombia; Amir Qaseem MD PhD, American College of Physicians, Philadelphia, Penn.; Richard M. Rosenfeld MD MPH, Department of Otolaryngology, SUNY Downstate Medical Center, Brooklyn, NY; Craig W. Robbins MD MPH, Center for Clinical Information Services, Kaiser Permanente Care Management Institute, Oakland, Calif.; Judith Thornton PhD, Centre for Clinical Practice, National Institute for Health and Clinical Excellence, Manchester, UK

Contributors: Holger Schünemann was the principal investigator. Holger Schünemann, Elie Akl and Wojtek Wiercioch conceptualized and designed the study with input from Nancy Santesso and Meghan McConnell. Meghan McConnell and Wojtek Wiercioch analyzed the data. Wojtek Wiercioch and Holger Schünemann drafted the manuscript. All of the authors contributed to the acquisition and interpretation of data, revised the manuscript critically for important intellectual content, approved the final version submitted for publication and agreed to be accountable for all aspects of the work.

Funding: This study was supported by internal funds from the McMaster GRADE centre, McMaster University. Kaja-Triin Laisaar receives institutional research support through grant IUT34-17 from the Ministry of Education and Research through the Estonian Research Council.

Data sharing: Data not shared.

Acknowledgements: The authors acknowledge Dr. John J. Riva for helping to pilot test the item-reduction survey and Dr. Gordon Guyatt for helping to field test the PANELVIEW tool. They also acknowledge all the participating guideline panel members who provided feedback for item generation and reduction, and helped with field testing of the tool.

Accepted: June 10, 2020

Correspondence to: Holger Schünemann, schuneh@mcmaster.ca