

A Comparative Analysis of Pulmonary and Critical Care Medicine Guideline Development Methodologies

Noah C. Schoenberg¹, Alan F. Barker², John Bernardo¹, Robin R. Deterding³, Jerrold J. Ellner¹, Dean R. Hess⁴, Neil R. MacIntyre⁵, Fernando J. Martinez⁶, and Kevin C. Wilson^{1,7}

¹Boston University, Boston, Massachusetts; ²Oregon Health & Science University, Portland, Oregon; ³University of Colorado, Denver, Colorado; ⁴Massachusetts General Hospital, Boston, Massachusetts; ⁵Duke University, Durham, North Carolina; ⁶Weill Cornell College of Medicine, New York, New York; and ⁷American Thoracic Society, New York, New York

Abstract

Rationale: The Institute of Medicine (IOM) standards for guideline development have had unintended negative consequences. A more efficient approach is desirable.

Objectives: To determine whether a modified Delphi process early during guideline development discriminates recommendations that should be informed by a systematic review from those that can be based upon expert consensus.

Methods: The same questions addressed by IOM-compliant pulmonary or critical care guidelines were addressed by expert panels using a modified Delphi process, termed the *Convergence of Opinion on Recommendations and Evidence* (CORE) process. The resulting recommendations were compared. Concordance of the course of action, strength of recommendation, and quality of evidence, as well as the duration of recommendation development, were measured.

Measurements and Main Results: When 50% agreement was required to make a recommendation, all questions yielded recommendations, and the recommended courses of action were 89.6% concordant. When 70% agreement was required, 17.9% of questions did not yield recommendations, but for those that did, the recommended courses of action were 98.2% concordant. The time to completion was shorter for the CORE process (median, 19.3 vs. 1,309.0 d; $P = 0.0002$).

Conclusions: We propose the CORE process as an early step in guideline creation. Questions for which 70% agreement on a recommendation cannot be achieved should go through an IOM-compliant process; however, questions for which 70% agreement on a recommendation can be achieved can be accepted, avoiding a lengthy systematic review.

Keywords: clinical practice guidelines; methodology; Institute of Medicine standards for trustworthy guidelines; Grading of Recommendations, Assessment, Development, and Evaluation approach

At a Glance Commentary

Scientific Knowledge on the Subject: Release of the Institute of Medicine standards for clinical practice guidelines heralded a new era in guideline development. However, the standards have had unintended negative consequences, including delays and high costs.

What This Study Adds to the Field: The introduction of a survey-based consensus approach, the *Convergence of Opinion on Recommendations and Evidence* process, as an early step in guideline creation can distinguish questions for which a systematic review is needed from those for which a systematic review is not needed, potentially saving time, effort, and money.

(Received in original form May 10, 2017; accepted in final form July 20, 2017)

Author Contributions: Conception and design: N.C.S. and K.C.W.; conduct of the study: N.C.S., A.F.B., J.B., R.R.D., J.J.E., D.R.H., N.R.M., F.J.M., and K.C.W.; data analysis: N.C.S. and K.C.W.; interpretation: N.C.S., A.F.B., J.B., R.R.D., J.J.E., D.R.H., N.R.M., F.J.M., and K.C.W.; drafting of the manuscript: N.C.S., A.F.B., J.B., R.R.D., J.J.E., D.R.H., N.R.M., F.J.M., and K.C.W.

Correspondence and requests for reprints should be addressed to Noah C. Schoenberg, M.D., The Pulmonary Center, R-304, 72 East Concord Street, Boston, MA 02118. E-mail: noah.schoenberg@bmc.org; or Kevin C. Wilson, M.D., The Pulmonary Center, R-304, 72 East Concord Street, Boston, MA 02118. E-mail: kwilson@thoracic.org or kcwilson@bu.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 196, Iss 5, pp 621–627, Sep 1, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201705-0926OC on July 21, 2017

Internet address: www.atsjournals.org

Release of the Institute of Medicine (IOM) standards for clinical practice guidelines in 2011 heralded a new era in guideline development (1). According to these standards, guidelines should be developed by panels with minimal conflicts of interest; a systematic review of the evidence should inform every recommendation; the rationale for each recommendation should be explicitly stated; recommendations should be articulated in a standardized fashion; and each recommendation should be rated according to its strength and the panel's confidence in the supporting evidence (i.e., quality of evidence) (1, 2). The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) method is a common approach used to articulate and rate recommendations (3, 4), but other systems also exist (5, 6).

Although well intentioned, the IOM standards have had unintended consequences, including delays and high costs, which are largely attributable to the requirement to perform a systematic review for each recommendation. Not only does performing this volume of systematic reviews dramatically increase the duration of a guideline's development, but the need to either outsource the systematic reviews or hire an experienced guideline methodologist to shepherd a guideline through development significantly increases the costs as well. As a result of these undesirable consequences, many guideline developers have reduced the scope of their guidelines or the number of guidelines that they produce, thus ultimately providing less guidance to the clinical community (unpublished program evaluation data, K. C. Wilson).

Proponents of the IOM approach argue that the more involved process is necessary to avoid recommending interventions that have no benefit or are harmful. Critics of the IOM approach argue that a panel of appropriately chosen experts is already well informed about the body of evidence and that therefore the systematic reviews are extraneous and serve only to create unnecessary delays. The critics further argue that there is a paucity of evidence that the systematic reviews demanded by the IOM approach produce more appropriate recommendations than expert panels do. There is no empiric evidence to support either perspective.

We hypothesize that the truth lies between the opposing views; specifically, we believe that some recommendations should be informed by systematic reviews, whereas others do not require a systematic review. It is our opinion that reducing the number of systematic reviews per guideline will decrease the duration and cost of guideline development, resulting in the dissemination of more guidance from experts to clinicians. The purpose of our study was to determine whether incorporation of an iterative, survey-based process modeled after the Delphi method (7) early during guideline development can discriminate recommendations that should be informed by a systematic review from those that can be based upon expert consensus alone. This study was approved by the Boston University Medical Campus Institutional Review Board (H-35497).

Methods

Survey Development

We identified eight American Thoracic Society (ATS)-sponsored clinical practice guidelines that had been completed but not yet presented publicly or published (8–15). These guidelines were created using an IOM standards-compliant process, meaning that an expert panel was assembled, questions were framed in the PICO (Population, Intervention, Comparator, and Outcome) format, full systematic reviews (i.e., multiple databases searched; studies selected and data extracted in duplicate) or pragmatic systematic reviews (i.e., only one database searched; studies selected and data extracted by a single individual) were performed for each question, and the quality of the evidence was rated using the GRADE approach (high, moderate, low, or very low quality). The systematic reviews were then used to inform the formulation of recommendations, and the GRADE approach was used to articulate the recommendations and to rate their strength (strong or conditional recommendation for or against the intervention).

For our study, the questions from each IOM process-derived guideline were composed into electronic multiple-choice surveys using SurveyMonkey software (SurveyMonkey, San Mateo, CA), retaining the format of the original guideline questions (see Figure E1A in the online

supplement). Each survey question consisted of four parts: (1) presentation of the PICO question, (2) a multiple-choice response asking for a strong or conditional recommendation for or against a course of action, (3) a second multiple-choice question asking participants to evaluate their impression of the quality of the evidence, and (4) a free-text box for comments.

Participants

For each guideline, a clinical expert who has expertise in the guideline's content area, but who was *not* involved in the original guideline's development, was identified by the senior author (K.C.W.). The clinical expert then identified 25–30 stakeholders, including physicians and nurses, who also have expertise in the guideline's content area. The stakeholders who were willing to participate in the study and were not involved in any aspect of development of the original guideline were formed into an expert panel. The conclusions and recommendations from the IOM process-derived guidelines were unknown to all participants.

Survey Administration

Electronic invitations were sent to each expert panel, inviting them to participate in the initial survey. Surveys were then administered via a modified Delphi method, which we termed the *Convergence of Opinion on Recommendations and Evidence* (CORE) process. This process consisted of administering the initial surveys to the expert panels, waiting a prespecified time (which varied from 7–14 d on the basis of participants' schedules), and then compiling the responses. These responses, which included answers to the multiple-choice questions as well as individual comments, were assembled and then provided to survey participants along with the original survey questions (Figure E1B). Participants were then asked again to respond to the survey questions after considering the results from the prior round. This process was repeated in an iterative fashion until stopping criteria were met. We prespecified our stopping criteria as either the completion of three rounds of surveys or the convergence of opinion as shown by more than 50% of respondents selecting a given choice, with the percentage of respondents selecting that option having increased from the prior round. (For

example, if 45% of respondents selected “conditional recommendation for the intervention” in round 1 and then 55% selected that same option during round 2, we would consider the stopping criteria met for that question.) If all questions in a given survey displayed convergence of opinion or three rounds were completed, we would halt the process. Figure 1 provides a graphical representation of the CORE process.

Outcomes Measures

Once the stopping criteria were met, the CORE-derived recommendations were tabulated using a 50% agreement threshold as the requirement for consensus. “Agreement” refers to the proportion of respondents who selected a given option on the final survey, “agreement threshold” refers to the extent of agreement needed for consensus, and “consensus” refers to achieving sufficient agreement to determine a recommendation. Thus, if at least 50% of respondents selected a strong or conditional recommendation for a given course of action in the final survey, then a recommendation was made for the course of action.

The CORE-derived recommendations were then compared with recommendations from the IOM standards-compliant guidelines. “Concordance” refers to the number of identical recommendations divided by the total number of recommendations. Thus, if 45 of the CORE-derived recommendations were identical to recommendations from the IOM standards-compliant guidelines and 5 were different, then concordance was 90%. We initially determined concordance of directionality (for or against the course of action); among recommendations with the same directionality, we also determined concordance of the strength (strong or conditional) and the quality of evidence (high, moderate, low, or very low).

To determine if requiring greater agreement would improve concordance, we performed a *post hoc* analysis using 60%, 70%, and 80% agreement thresholds as requirements for consensus for the CORE-derived recommendations. As an example, if 75% of respondents selected a strong or conditional recommendation for a course of action in the final survey, then recommendations for the course of action could be made when a 50%, 60%, or 70%

agreement threshold was required for consensus, but no recommendation could be made when an 80% agreement threshold was required. Concordance was determined for each agreement threshold.

Additionally, we compared durations from question completion to recommendations for the CORE and the IOM standards-compliant processes. Durations of the CORE process were measured from initial survey distribution to final recommendations; durations of the IOM standards-compliant process were estimated on the basis of the interval from guideline project commencement to manuscript submission, with an assumed average of 9 months subtracted for panel composition (including conflict-of-interest screening), PICO question determination, and manuscript writing.

Statistical Analysis

Statistical analysis was performed using SAS version 9.32 software (SAS Institute, Cary, NC). For the 50% and 70% agreement thresholds for consensus, an agreement kappa value was calculated between the IOM standard process and the CORE process (16). Duration differences were examined using a Wilcoxon rank-sum test, and a Hodges-Lehmann estimator provided a median and 95% confidence limits.

Results

Participation

Initial survey response rates ranged from 76 to 100%, and they ranged from 72 to 100% on subsequent rounds. All surveys had a minimum of 15 responses in the final round. Results were calculated on the basis of responses provided in the final round (responses of participants who completed only early surveys were not included). This maintained the integrity of the iterative Delphi approach.

Agreement Thresholds for Consensus

At the prespecified 50% agreement threshold for consensus, recommendations could be made for all questions. At higher agreement thresholds, there were questions for which a recommendation could not be made, because the response percentages did not exceed the agreement threshold. For these questions, we stated that no recommendation could be given. As the agreement threshold was raised, the number

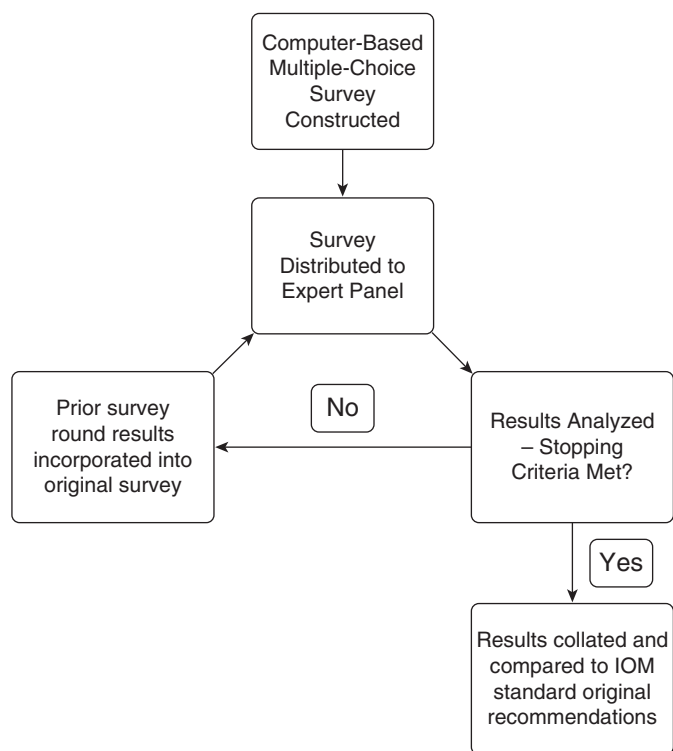


Figure 1. The Convergence of Opinion on Recommendations and Evidence process. IOM = Institute of Medicine.

of questions for which recommendations could be made decreased; specifically, at agreement thresholds of 50%, 60%, 70%, and 80%, the percentages of questions for which a recommendation could be made were 100%, 94%, 82%, and 76%, respectively (Table 1).

Directionality of recommendations.

When the CORE process was conducted using a 50% agreement threshold for consensus, 60 (89.6%) of 67 questions yielded concordant recommendations for directionality and 7 (10.4%) of 67 yielded discordant recommendations. This corresponds to a kappa value of 0.71 with an SE_{κ} of 0.104. When the agreement threshold for consensus was increased to 70%, fewer recommendations were made, but the recommendations more closely mirrored the IOM process-derived recommendations. Specifically, 12 (17.9%) of 67 questions did not reach sufficient agreement to yield a recommendation; however, among the 55 remaining recommendations, 54 (98.2%) were concordant with the IOM process-derived recommendations, and only 1 (1.8%) was discordant (Table 1). This corresponds to a kappa value of 0.941 with an SE_{κ} of 0.058, which, depending on the criteria employed, rates as either “almost perfect” or “excellent” agreement (16). Among the 12 questions for which there was insufficient agreement to yield a recommendation at the 70% agreement threshold, 6 (50%) had yielded discordant recommendations and 6 (50%) had yielded concordant recommendations at the 50% agreement threshold.

Most discordant recommendations occurred in the context of low- or very low-quality evidence, as defined by the IOM standards-compliant process. At the 50% agreement threshold, 5 (71.4%) of 7 discordant recommendations were associated with low- or very low-quality

evidence. At the 70% agreement threshold, the lone discordant recommendation was associated with low-quality evidence.

Strength of recommendations and quality of evidence. Among the 60 recommendations with directional concordance that were derived using a 50% agreement threshold, 43 (71.7%) had concordance of the strength of the recommendation and 21 (35%) had concordance of the quality of evidence. Of the 17 recommendations for which the strength differed, the CORE process overestimated the strength of 12 (71%) and underestimated 5 (29%). Of the 39 recommendations where the quality-of-evidence evaluation differed, the CORE process overestimated the quality of evidence of 36 (92.3%) while underestimating only 3 (7.7%) (Table 2). At the 70% agreement threshold, 29 (53.7%) of 54 of the recommendations displayed concordance of strength, and 10 (18.5%) of 54 of the recommendations had concordance of the quality of evidence (Table 2). Among discordant results at the 70% agreement threshold, the CORE process overestimated both the strength of the recommendations and the quality of evidence.

Duration

The median time to completion was significantly shorter for the CORE process than for the IOM standards-compliant process (median, 19.3 d vs. 1,309 d; $P = 0.0002$; interval midpoint, 1,181 d; 95% confidence interval, 711–1,651 d).

Discussion

This study was conceived with the goal of creating a process to streamline guideline creation while maintaining recommendation validity. It was designed as

a proof-of-concept study to explore whether expert knowledge of a subject area assessed through a modified Delphi process was sufficiently high to replace portions of the systematic review-based, IOM standards-compliant process. We found that using a 50% agreement threshold for consensus resulted in an approximately 90% concordance rate with regard to directionality of the recommendation. We felt that this was unacceptably low. When we increased the agreement threshold for consensus to 70%, there were 12 questions (17.9%) for which no recommendation could be made; however, for the remaining 55 questions (82.1%) for which a recommendation was possible, the concordance rate increased to 98%, with only one discordant recommendation. We consider this to be an excellent concordance rate, especially considering that no descriptions of the evidence were presented to the panel prior to generating these recommendations. Also critical is the fact that the CORE process produced concordant recommendations to the IOM standards-compliant process in a fraction of the time.

Notably, however, concordance for strength of recommendation and quality of evidence was moderate to poor. Participants were not provided any formal training on how to rate quality of evidence or strength of recommendations as part of this study. Supporting this as a contributor to the moderate to poor concordance, a previous study reported interrater reliability coefficients ranging from 0.49 to 0.84 following two training sessions on rating the quality of evidence using GRADE (17); however, participants were health research methodology students and GRADE working group members rather than guideline panelists. The low concordance rates in our study reflected the tendency of experts to overestimate the quality of

Table 1. Concordance of the Convergence of Opinion on Recommendations and Evidence-derived Recommendations for Directionality

	Consensus Threshold			
	50%	60%	70%	80%
Recommendations made (% among total questions)	67 (100.0%)	63 (94.0%)	55 (82.1%)	51 (76.1%)
Concordant recommendations (% among recommendations made)	60 (89.6%)	58 (92.1%)	54 (98.2%)	50 (98.0%)
Discordant recommendations (% among recommendations made)	7 (10.4%)	5 (7.9%)	1 (1.9%)	1 (2.0%)

Table 2. Concordance of the Convergence of Opinion on Recommendations and Evidence-derived Recommendations for Strength of Recommendations and Quality of Evidence

	Consensus Threshold	
	50%	70%*
Strength of recommendation		
Concordant strength (% among recommendations with directional concordance)	43 (71.7%)	29 (53.7%)
Discordant strength (% among recommendations with directional concordance)	17 (28.3%)	12 (22.2%)
Overestimated (% among recommendations with discordant strength)	12 (70.5%)	9 (75.0%)
Underestimated (% among recommendations with discordant strength)	5 (29.4%)	3 (25.0%)
Quality of evidence		
Concordant QoE (% among recommendations with directional concordance)	21 (35.0%)	10 (18.5%)
Discordant QoE (% among recommendations with directional concordance)	39 (65.0%)	17 (31.5%)
Overestimated (% among recommendations with discordant QoE)	36 (92.3%)	17 (100.0%)
Underestimated (% among recommendations with discordant QoE)	3 (7.7%)	0 (0.0%)

Definition of abbreviation: QoE = quality of evidence.

*At the 70% consensus threshold, no recommendations could be made for strength and quality of evidence for 13 (24.1%) of 54 and 27 (50.0%) of 54, respectively.

available evidence. This overestimation of the quality of evidence would then naturally lead to a tendency to emphasize a course of action by increasing the strength of the recommendation. One goal of a future prospective study evaluating the CORE process would be to provide formal guidance in evaluation of the quality of evidence and then examine the responses to see if this improves the concordance with evaluations derived via the IOM standards-compliant process.

Although the concordance on the quality of the evidence was suboptimal, the external validity of the GRADE ranking of the quality of evidence is itself in question, owing to its unclear ability to predict static estimates of effect (18, 19). Given this inherent uncertainty in the accuracy of the quality of evidence evaluations, a lack of agreement with this ranking is less problematic for the CORE process than if the most important aspect of the recommendation—the directionality—had lacked agreement. Given that one of the major goals of the IOM standards is to improve transparency, one might argue that the poor concordance indicates that the IOM standards should require reporting the level of agreement among panel members for rating the quality of evidence

and the strength of the recommendation. Although we appreciate the idea, we question its feasibility for several reasons. First, formal votes (by which an agreement level would be obtained) do not always occur during guideline development, because recommendations are often based upon consensus via discussion. Second, even if such votes did regularly occur, the relatively small size of most guideline development panels would make calculating an agreement level (kappa) for quality of evidence or strength of recommendation difficult.

Upon closer examination, the majority of the discordant recommendations (directionality) occurred in the setting of low or very low quality of evidence. This suggests that discordant recommendations occur primarily in subject areas for which there is limited evidence or that tend to be more controversial. By raising the agreement threshold to 70%, we were able to optimize the number of recommendations given while still eliminating those either too controversial or too lacking in evidence to simply rely on expert opinion to generate recommendations.

Thus, we propose that the CORE process be performed as an initial step in guideline creation, following panel

composition and question finalization (Figure 2). Questions for which 70% agreement on a recommendation cannot be achieved should go through an IOM standards-compliant process, ensuring that questions that are controversial or uncertain are systematically explored. Recommendations for which 70% agreement can be achieved through the CORE process can be accepted, thereby avoiding a lengthy and costly systematic review. Recommendations that are accepted on the basis of surpassing the agreement threshold would not necessarily have ratings of strength of the recommendation or quality of evidence attached, unless we are able to demonstrate in future studies that formal training improves the ability of the expert panel to accurately assess the quality of evidence. With this proposed process, we expect a shorter duration of guideline development, which would result in cost savings; however, this could not be confirmed in the present study because of incomplete information on the final costs of the IOM standards-compliant guidelines. We expect that decreasing the cost of guideline development and lessening the time required will simultaneously permit guidelines to be developed with a greater scope and enable expert guidance to reach the clinician sooner.

Our study has several important limitations. First, our study could not include the same experts as were included in the original guideline panels, and thus it is possible that different expert panel compositions rather than the different processes were the cause of the discordant recommendations. Future studies should use the same panel to compare recommendations derived from the CORE process with those derived using an IOM standards-compliant approach. Second, our guideline questions were confined to pulmonary and critical care medicine (the content area of the clinical practice guidelines used as test cases), so generalizability beyond those fields is uncertain. Further research is required to demonstrate applicability to other medical specialties. Third, our estimations of duration of the IOM standards-compliant process are merely estimates because the exact amount of time required for panel composition, conflict-of-interest vetting, PICO question development, and manuscript writing was presumed to be approximately 9 months. In reality, these

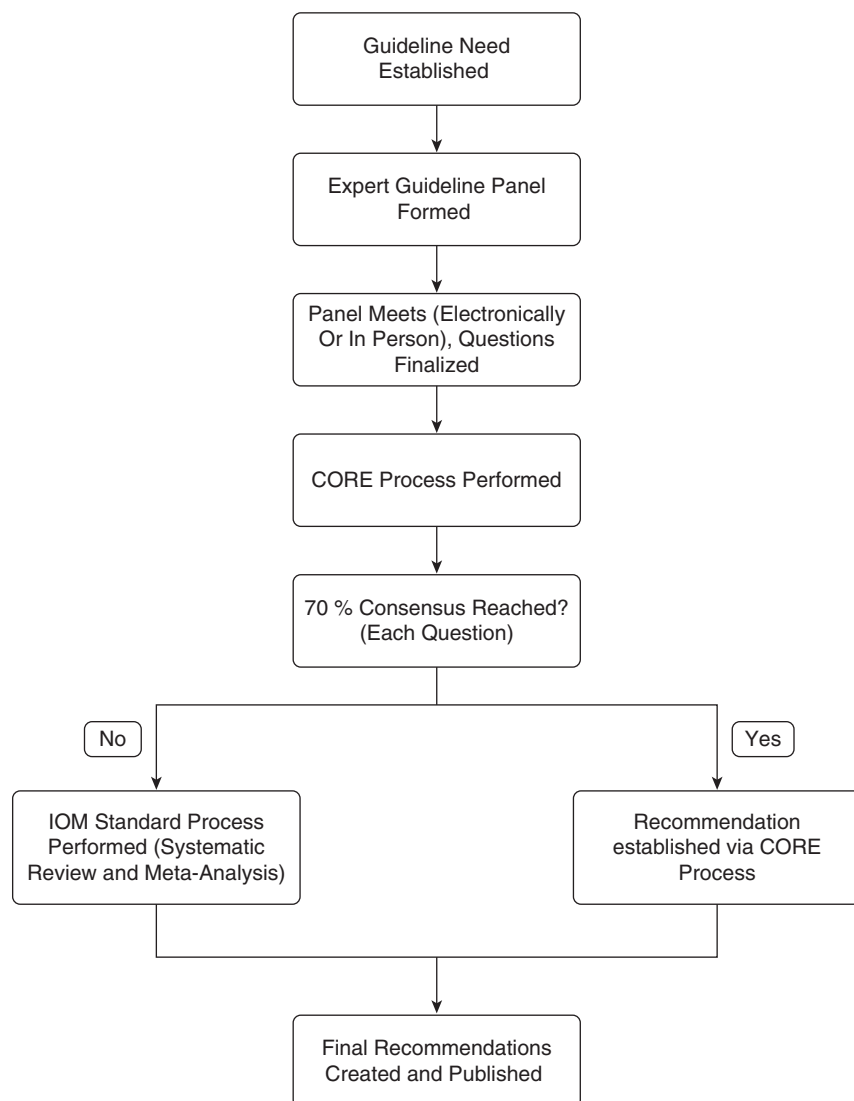


Figure 2. Proposed Convergence of Opinion on Recommendations and Evidence (CORE) process utilization algorithm. IOM = Institute of Medicine.

tasks may have taken more or less time. Finally, our stopping criteria were arbitrary, and the impact of those criteria is unknown. By adjusting those criteria, one could potentially fine-tune the CORE process to increase its accuracy and improve its utility.

Thus, this trial serves as a proof-of-concept study evaluating an alternative guideline construction process. We found that the CORE process, a modified Delphi approach based on iterative surveys and expert knowledge, accurately reproduced a high percentage of guideline recommendations developed through an IOM standards-compliant process. Though further studies are needed to confirm these results and fine-tune the CORE process, we

propose that it serve as a preliminary step in the guideline construction process, thus significantly decreasing the number of systematic reviews required to create a guideline. This in turn would decrease the time and cost required, facilitating the delivery of more and broader guidelines into the hands of clinicians. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Drs. Allan Walkey, Elizabeth Klings, Renda Wiener, and George O'Connor for statistical guidance and editing assistance. Additionally, the authors thank the following participants in our Convergence of Opinion on Recommendations and Evidence (CORE) processes who

generously gave their time and effort in the interest of trying to improve the efficiency of guideline development. *American Thoracic Society (ATS) Guidelines on Diagnosis of Pediatric Recurrent or Persistent Wheezing* (8): Robin Deterding (leader), Ibrahim Ahmed Janahi, Len Bacharier, R. Paul Boesch, Andy Bush, Mateja Cernelc-Kohan, Cori Daines, Sharon Dell, Jonathan Gaffin, Theresa Guilbert, Adam Jaffe, Carolyn Kercksmar, Nadia Krupp, Paul E. Moore, Wayne Morgan, Mary Nevin, Sande Okelo, Jonathan Popler, Deepa Rastogi, Kristie Ross, and Peter Sly. *ATS/CDC/Infectious Diseases Society of America (IDSA) Guidelines on the Treatment of Drug-susceptible Tuberculosis* (9): John Bernardo (leader), Marcos Burgos, Jon Warkentin, Jenny Flood, Diana Nilsen, Elizabeth Talbot, Dean Tsukayama, Gerald Mazurek, John Jereb, Lynn Sosa, Shu Wang, Jason Stout, David Cohn, Randall Reves, Maureen Murphy-Weiss, Barbarah Martinez, and Diana Fortune. *ATS/Japanese Respiratory Society (JRS) Guidelines on Diagnosis and Treatment of Lymphangioleiomyomatosis* (10): Alan Barker (leader), Steve Ruoss, George Pappas, Adrian Shifren, Allan Glanville, Souheil El-Chemaly, Matt Drake, Vera Krymskaya, Toshinori Takada, Tony Eissa, Maryl Kreider, Charles Burger, Augustine Lee, Richard Helmers, Joseph Parambil, Kevin Flaherty, Rosa Estrada y Martin, Marilyn Glassberg, Gordon Yung, Monica Goldklang, Richard Lubman, Kyle Brownback, and Teng Moua. *ATS/American College of Chest Physicians (CHEST) Guidelines on Liberation from Mechanical Ventilation* (11): Dean Hess (leader), Nick Hill, R. Scott Harris, Stefano Nava, Paolo Navalesi, Jordi Mancebo, Richard Branson, Robert Kacmarek, Richard Kallet, Neil MacIntyre, Gerald Criner, Lluís Blanch, Luca Bigatello, William Hurford, Suhail Raof, David Kaufman, Steven Holets, Richard Oeckler, Steven Deem, Marco Ranieri, Younsuck Koh, Niall Ferguson, David Berlin, Daniel Talmor, Rolf Hubmayr, and Brian Kavanagh. *ATS/CDC/IDSA Guidelines on the Diagnosis of Tuberculosis* (12): Jerrold Ellner (leader), Elizabeth Talbot, Robert Horsburgh, Claudia Denkinger, Mark Nicol, Rinn Song, Max O'Donnell, Robert Husson, Kevin Fennelly, Pennan Barry, Andy Vernon, Neil Schluger, Adithya Cattamanchi, Fred Gordin, Rocco Hurtado, Julie Higashi, Christina Ho, Masahiro Narita, G.B. Migliori, Francisco Blasi, and Giovanni Sotgiu. *ATS/Society of Critical Care Medicine (SCCM)/European Society of Intensive Care Medicine (ESICM) Guidelines on Mechanical Ventilation in Acute Respiratory Distress Syndrome* (13): Neil MacIntyre (leader), Greg Schmidt, Eric Osborne, David Bowton, Craig Rackley, Christopher Cox, Joseph Govert, Jonathan Truweit, Carl Shanholtz, Scott Epstein, John Hollingsworth, Stephanie Levine, Lisa Moores, Robert Chatburn, Josh Benditt, Rick Albert, Antonio Anzueto, Paolo Pelosi, Andres Esteban, Jay Johannigman, Carolyn Calfee, Eduardo Mireles-Cabadevilla, Tim Girard, and Michael Sjoding. *European Respiratory Society (ERS)/ATS Guidelines on the Treatment of Chronic Obstructive Pulmonary Disease (COPD) Exacerbations* (14): Fernando Martinez (leader), Shawn Aaron, Nicolino Ambrosino, David Au, Richard Casaburi, Bart Celli, Enrico

Clini, Mark Dransfield, Leo Fabbri, Peter Lindenauer, MeilAn Han, Francois Maltais, Dennis Niewoehner, Nicolas Roche, Carly Rochester, Frank Scierba, Sanjay Sethi, Don Sin, Daiana Stolz, Robert Stockley, James Stoller, Charlie Strange, Geert Verleden, Claus

Vogelmeier, and Richard ZuWallack. *ERS/ATS Guidelines on the Prevention of COPD Exacerbations* (15): Fernando Martinez (leader), Shawn Aaron, Nicolino Ambrosino, David Au, Richard Casaburi, Bart Celli, Enrico Clini, Mark Dransfield, Leo Fabbri, Peter Lindenauer,

MeiLan Han, Francois Maltais, Dennis Niewoehner, Nicolas Roche, Carly Rochester, Frank Scierba, Sanjay Sethi, Don Sin, Daiana Stolz, Robert Stockley, James Stoller, Charlie Strange, Geert Verleden, and Claus Vogelmeier.

References

- Institute of Medicine. Clinical practice guidelines we can trust: standards for developing trustworthy clinical practice guidelines (CPGs). Washington, DC: Institute of Medicine; 2011 [accessed 2016 Oct 10]. Available from: <http://www.nationalacademies.org/hmd/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx>
- Qaseem A. A perspective on the Guidelines International Network and the Institute of Medicine's Proposed Standards for Guideline Development. Rockville, MD: National Guideline Clearinghouse, Agency for Healthcare Research and Quality; 2013 [accessed 2017 Jul 31]. Available from: <https://www.guideline.gov/expert/expert-commentary/43913/aperspective-on-the-guidelines-international-network-and-the-institute-of-medicines-proposedstandards-for-guideline-development#>
- Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, et al.; ATS Documents Development and Implementation Committee. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006;174:605–614.
- Guyatt G, Akl EA, Oxman A, Wilson K, Puhan MA, Wilt T, Gutterman D, Woodhead M, Antman EM, Schünemann HJ; ATS/ERS Ad Hoc Committee on Integrating and Coordinating Efforts in COPD Guideline Development. Synthesis, grading, and presentation of evidence in guidelines: article 7 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proc Am Thorac Soc* 2012;9:256–261.
- Jacobs AK, Kushner FG, Ettinger SM, Guyton RA, Anderson JL, Ohman EM, Albert NM, Antman EM, Arnett DK, Bertollet M, et al. ACCF/AHA clinical practice guideline methodology summit report: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2013;127:268–310.
- U.S. Preventive Services Task Force. Grade definitions [accessed 2017 Feb 12]. Available from: <https://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions>
- Mead D, Moseley L. The use of Delphi as a research approach. *Nurse Res* 2001;8:4–23.
- Ren CL, Esther CR Jr, Debley JS, Sockrider M, Yilmaz O, Amin N, Bazy-Asaad A, Davis SD, Durand M, Ewig JM, et al.; ATS Ad Hoc Committee on Infants with Recurrent or Persistent Wheezing. Official American Thoracic Society Clinical Practice Guidelines: diagnostic evaluation of infants with recurrent or persistent wheezing. *Am J Respir Crit Care Med* 2016;194:356–373.
- Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, Chaisson LH, Chaisson RE, Daley CL, Grzemska M, et al. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: treatment of drug-susceptible tuberculosis. *Clin Infect Dis* 2016;63:e147–e195.
- McCormack FX, Gupta N, Finlay GR, Young LR, Taveira-DaSilva AM, Glasgow CG, Steagall WK, Johnson SR, Sahn SA, Ryu JH, et al.; ATS/JRS Committee on Lymphangioleiomyomatosis. Official American Thoracic Society/Japanese Respiratory Society Clinical Practice Guidelines: lymphangioleiomyomatosis diagnosis and management. *Am J Respir Crit Care Med* 2016;194:748–761.
- Schmidt GA, Girard TD, Kress JP, Morris PE, Ouellette DR, Alhazzani W, Burns SM, Epstein SK, Esteban A, Fan E, et al.; ATS/CHEST Ad Hoc Committee on Liberation from Mechanical Ventilation in Adults. Official Executive Summary of an American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: liberation from mechanical ventilation in critically ill adults. *Am J Respir Crit Care Med* 2017;195:115–119.
- Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, Keane J, Lewinsohn DA, Loeffler AM, Mazurek GH, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: diagnosis of tuberculosis in adults and children. *Clin Infect Dis* 2017;64:e1–e33.
- Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, Adhikari NKJ, Amato MBP, Branson R, Brower RG, et al.; American Thoracic Society, European Society of Intensive Care Medicine, and Society of Critical Care Medicine. An Official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2017;195:1253–1263.
- Wedzicha JA, Miravittles M, Hurst JR, Calverley PM, Albert RK, Anzueto A, Criner GJ, Papi A, Rabe KF, Rigau D, et al. Management of COPD exacerbations: a European Respiratory Society/American Thoracic Society guideline. *Eur Respir J* 2017;49:1600791.
- Wedzicha JA, Calverley PMA, Albert RK, et al. Prevention of COPD exacerbations: a European Respiratory Society/American Thoracic Society Guideline. *Eur Respir J* (In press).
- Steiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Vol. 2. Oxford, UK: Oxford University Press; 2008.
- Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, Kulasegaram M, Christensen R, Guyatt GH, Falck-Ytter Y, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66:736–742, quiz 742.e1–742.e5.
- Norris SL, Bero L. GRADE methods for guideline development: time to evolve? *Ann Intern Med* 2016;165:810–811.
- Gartlehner G, Dobrescu A, Evans TS, Bann C, Robinson KA, Reston J, Thaler K, Skelly A, Glechner A, Peterson K, et al. The predictive validity of quality of evidence grades for the stability of effect estimates was low: a meta-epidemiological study. *J Clin Epidemiol* 2016;70:52–60.