
Michael S Broder, Sarah N Gibbs, Irina Yermilov

Outcomes Research, Partnership for Health Analytic Research (PHAR), LLC, Beverly Hills, CA, USA

Correspondence: Michael S Broder, Partnership for Health Analytic Research (PHAR), LLC, 280 S Beverly Drive, Suite 404, Beverly Hills, CA, 90212, USA, Tel +1-310-858-9555, Fax +1-310-858-9550, Email mbroder@pharllc.com

Abstract: The RAND/UCLA modified Delphi panel method is a formal group consensus process that systematically and quantitatively combines expert opinion and evidence by asking panelists to rate, discuss, then re-rate items. The method has been used to develop medical society guidelines, other clinical practice guidelines, disease classification systems, research agendas, and quality improvement interventions. Traditionally, a group of experts meet in person to discuss results of a first-round survey. After the meeting, experts complete a second-round survey used to develop areas of consensus. During the COVID-19 pandemic, this aspect of the method was not possible. As such, we have adapted the method to conduct virtual RAND/UCLA modified Delphi panels. In this study, we present a targeted literature review to describe and summarize the existing evidence on the RAND/UCLA modified Delphi panel method and outline our adaptation for conducting these panels virtually. Transitioning from in-person to virtual meetings was not without challenges, but there have also been unexpected advantages. The method we describe here can be a cost-effective and efficient alternative for researchers and clinicians.

Keywords: Delphi panel, expert panel, consensus, virtual meeting, COVID-19, pandemic

Introduction

The RAND/UCLA modified Delphi panel method is a formal group consensus process that systematically and quantitatively combines expert opinion and evidence by asking panelists to rate, discuss, then re-rate items. In brief, the steps include a literature review, selection of panelists, generation of a rating form, a first-round rating form survey, an in-person meeting where panelists discuss areas of disagreement, final ratings and analysis of those ratings, and the development of a written summary of areas of agreement.

Such panels have been used to develop medical society guidelines, other practice guidelines, disease classification systems, research agendas, and quality improvement interventions. Guidelines developed using this method have content, construct, and predictive validity. Results of modified Delphi panels conducted using the same evidence base produce similar results, and patients treated according to the resulting guidelines have been shown to have improved outcomes.

The panel meeting, a significant component of the method where panelists discuss areas of disagreement and the rationale for their first-round ratings, is traditionally held in person. During the COVID-19 pandemic, meetings could not be held in person. However, to maintain the integrity of the method while continuing to develop practice guidelines during the ongoing pandemic, we adapted this method and successfully conducted several virtual RAND/UCLA modified Delphi panels since March 2020. In this study, we conducted a targeted literature review to describe and summarize the existing evidence on the RAND/UCLA modified Delphi panel method and outlined our adaptation for conducting these panels virtually.

History of the RAND/UCLA Delphi Panel Method

The “Delphi method” was originally developed by the RAND Corporation in the 1950s as a way to obtain group consensus on military decisions. The technique was designed to avoid direct confrontation among experts by not completing ratings in
person and instead relied on repeated, individual questioning of experts through surveys, later called “rating forms.” In this way, the method eliminated “specious persuasion,” where the participant with the strongest convictions or greatest supposed authority pushes other participants to agree against their own judgment. The method used successive rounds of surveys. Expert respondents were shown the group medians after each round and asked to explain their reasoning if their ratings fell outside a certain range. Several rounds were conducted in this way until “consensus” emerged.

In the 1980s, RAND adapted the method in partnership with UCLA for use in the medical setting (becoming the RAND/UCLA modified Delphi panel method). With this adaptation, researchers added the development of a summary of relevant literature to provide uniform context to experts and limited the number of repeated individual questionnaires to two (rather than an unlimited number), one completed before and one after an in-person meeting. Researchers found that when addressing clinical questions, a discussion was vital to debating areas of disagreement. However, as in the original design, the group was never required to agree during the meeting. Instead, the second-round ratings were used as a summary of the group consensus (defined mathematically based on the number of low versus high ratings).

**Contemporary Use of the RAND/UCLA Modified Delphi Panel Method**

The RAND/UCLA modified Delphi panel method has been used to develop clinical practice guidelines in a variety of areas. For example, it has been used to determine the appropriateness of and treatment preference for coronary angiography, percutaneous transluminal coronary angioplasty, and coronary artery bypass grafting. It has been used to develop criteria for the appropriate use of urinary catheters in hospitalized patients, peripherally inserted central catheters, and transfusions in heptectomies. It has also been used to identify circumstances when spinal mobilization and manipulation are inappropriate for patients with chronic low back pain and when and how to screen for neurotrophic keratopathy. Guidelines have been developed using the method to enhance the use of medical and surgical measures for recurrent stroke prevention, to identify systemic treatments for unresectable metastatic well-differentiated carcinoid tumors, and to determine the appropriateness of various biologics for Crohn’s disease. The method has helped determine adequate follow up intervals for patients with Cushing’s Disease at different phases in their treatment course and to describe specific circumstances when it is appropriate to consider tapering thrombopoietin receptor agonists in primary immune thrombocytopenia.

Why the Method Works

Multiple aspects of the RAND/UCLA modified Delphi panel method combine to make it a successful tool for developing clinical guidelines. First, the structure of the rating form focuses decision-making. The form includes hundreds of patient scenarios differing across clinically relevant characteristics. It is typically structured as a grid, in which scenarios in nearby cells differ only on one or two characteristics. This design encourages panelists to think granularly about whether each characteristic alone has an impact on their ratings. In clinical practice, physicians faced with complex and diverse patients often make intuitive judgments using what has been called “System 1” decision-making—a faster, unconscious, and less intentional decision-making process. While medical students are trained to solve clinical problems by separately considering multiple aspects of the clinical picture, expert physicians typically use the more intuitive process. However, medical decision-making relying only on System 1 thinking may be more prone to error due to cognitive biases. For example, in clinical practice, when deciding if an intervention is appropriate for an older patient with a complex set of risk factors, a clinician might focus only on age—substituting the simpler question for a more complex one. The detailed ratings used in the Delphi panel rating form encourage panelists to use the slower, more logical, and conscious “System 2” thinking. With each risk factor explicitly listed, they cannot easily be ignored.

The use of two rounds of ratings with group discussion interposed is also likely a factor in the success of the process. It has long been observed that the accuracy of an estimate can be improved by combining estimates from different individuals, sometimes called the “wisdom of the crowd.” Clinical decisions often hinge on the likelihood of a particular outcome. For example, the likelihood of myocardial infarction in an untreated patient with cardiac risk factors drives the decision to prescribe a statin. Panelists asked to rate the appropriateness of such a prescription are implicitly estimating the risk of doing so. Combining individual panelist ratings should improve the accuracy of this estimation.
More recently, researchers have demonstrated “the wisdom of the inner crowd,” finding that aggregated estimates all made by a single individual may be more accurate than that individual’s “best guess.” When an individual’s second estimate is constructed from the perspective of someone with whom they disagree, accuracy improves even more. During the panel discussion, the moderator guides panelists to share their reasoning on scenarios for which there was disagreement. Panels purposely include experts from different backgrounds (eg specialty, practice type, geographic regions). Explicit discussion of differences of opinion, where panelists must consider the opinion of others with views different than their own, may lead the panelists to more consistent and logical positions. The increased agreement seen in almost all second-round Delphi panel ratings is consistent with this observation. In the second round, each individual panelist re-rates the items, having now considered not only what someone they disagree with (eg another panelist) thinks, but why she thinks that way.

Beyond the cognitive process noted above, the discussion at the in-person meeting also results in decreased disagreement among the group by harmonizing panelists’ interpretations of definitions used in the rating form. Panelists discuss definitions, clarify items that were unclear, and come to a common understanding of terms that can then be used in the resulting clinical guidance. For example, in a panel to determine the appropriateness of tapering thrombopoietin receptor agonists in primary immune thrombocytopenia, characteristics that could not be defined numerically (eg major/minor bleeding) versus those that could (eg platelet count), benefited from group discussion that made it more likely that all panelists were interpreting the characteristic in the same way.

Lastly, by having panelists complete the rating form independently a second time following the meeting rather than requiring the group to agree at the meeting, helps to prevent the discussion from stalling if panelists refuse to agree. In The Wisdom of the Crowds, Surowiecki argues that too much communication from large groups can be unmanageable and inefficient. The structured nature of the Delphi panel meeting discussion, in which the moderator works through ratings completed in advance, limits these inefficiencies. If arguments arise, the moderator can simply move to the next item, acknowledging that there may be some remaining areas of disagreement in the second-round results.

Validity and Reliability of the Method
By avoiding certain cognitive biases and harnessing the collection knowledge of experts without the limitations of large crowds, guidelines developed using this method have content, construct, and predictive validity. Using a retrospective medical chart review design, researchers found that patients with heart disease eligible for revascularization and identified as likely to benefit from surgery per guidelines developed using the RAND/UCLA modified Delphi panel process, had better outcomes (lower mortality and prevalence of angina) after revascularization than did patients treated medically. This study was repeated prospectively at three hospitals in London, as well as through a post-hoc analysis of clinical trial data, with similar results. Recently, a group of heart disease specialty societies (eg the American College of Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons, and the American Association for Thoracic Surgery) updated these revascularization guidelines using the same method.

In addition, the method has been found to be reliable. Studies have reported a test-retest reliability of >0.9 using the same panelists 6–8 months later and kappa statistics across several panels with different members similar to those of some common diagnostic tests. Independent panels using this method also produce similar ratings to one another, although the degree of similarity depends on the level of evidence available. A review of Delphi panels showed 90% agreement among the panels that used randomized control trial evidence compared to 70–80% agreement in the panels which used a weaker evidence base.

Adapting the RAND/UCLA Modified Delphi Panel Method to a Virtual Setting
The RAND/UCLA modified Delphi panel method places emphasis on the in-person discussion. However, during a time when in-person meetings could no longer take place due to the COVID-19 pandemic, we adapted the discussion to a virtual setting (Figure 1).
Certain aspects of the method did not need to change. For example, we continued to include 10–13 panelists and were able to discuss a similar range of items (200–700) (Table 1). The comprehensive summary of relevant literature typically developed by the research team had often already been shared electronically. Instead of sharing printed studies at the in-person meeting, we provided an online link to a folder with the electronic publications cited in the literature review. To inform the development of the rating form, researchers had previously conducted and continued to conduct a series of phone conversations with all participating panelists. During these conversations, the research team identifies characteristics critical to clinician decision-making on the panel topic and uses this information to develop the survey. Lastly, prior to the meeting in any format, panelists had been asked to complete their ratings electronically. These first-round ratings allow the moderator to identify patterns of agreement and disagreement and prepare for the meeting.

Instead of meeting in person, we adapted the method by organizing and moderating the panel meetings virtually through video conference. In person, meetings were held over 1–2 full days in a central geographic location. We believed virtual meetings of this length would be unproductive. Instead, we redesigned the discussion into several 2- to 4-hour blocks of time on consecutive days. We have been able to complete these guided discussions in a shorter amount of time virtually (~6–7 hours logged into a meeting) than was required when meeting in person (1–2 full days including travel to attend a 6–9-hour meeting) (Table 1). In addition, shorter virtual segments allow participating experts, usually practicing physicians, to maintain some parts of their busy clinic schedules on these days. As a result, scheduling these meeting times has been simpler and allowed us to engage a more diverse group of experts, including those who do not primarily conduct research or those who were reluctant to travel for a panel meeting, even prior to the pandemic.
During our first meetings at the start of the pandemic, we worked with our panelists ahead of the meeting to familiarize them with the video conferencing technology to avoid technological challenges on the day of the meeting. As the pandemic evolved, our panelists became more adept at video conferencing and have not needed this additional step.

There are some limitations to this virtual panel method. First, in person, meetings often generated an informal interchange among panelists during breaks or meals, which often built rapport and respect among the group. We have not been able to recreate this virtually. Second, international panels with experts from across the globe, though less expensive and easier to plan due to lack of travel, are harder to schedule given time differences. It can be difficult to find several hour segments of time during which all panelists are willing to engage in a virtual discussion.

Nevertheless, in several virtual panels since early 2020, we have successfully reduced disagreement and achieved consensus on a variety of clinical topics (Table 1). For example, during a panel held on March 18–19, 2020, the proportion of items with disagreement decreased by half following the panel meeting. Upcoming publications will discuss other modified Delphi panel findings in detail.

**Discussion**

In response to the COVID-19 pandemic, we adapted the RAND/UCLA modified Delphi panel method to include a virtual, rather than in person, meeting. The result is a cost-effective and efficient alternative. Virtual panel meetings include the same elements that have made the method a success (eg a diverse group of 9–13 panelists, a reduction in

<table>
<thead>
<tr>
<th>Project Title</th>
<th>In Person / Virtual</th>
<th>Number of Panelists</th>
<th>Length of Meeting</th>
<th>Number of Items Rated</th>
<th>Percent Disagreement from First-Round to Second-Round Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of On-Demand Treatments for OFF Episodes in Parkinson's Disease: Guidance from a RAND/UCLA Modified Delphi Consensus Panel</td>
<td>Virtual</td>
<td>12 (5 female, 7 male)</td>
<td>7 hours</td>
<td>432</td>
<td>67, 32</td>
</tr>
<tr>
<td>Expert Consensus on the Testing and Medical Management of PIK3CA-Related Overgrowth Spectrum</td>
<td>Virtual</td>
<td>13 (7 female, 6 male)</td>
<td>7 hours</td>
<td>217 (first-round), 115 (second-round)</td>
<td>19, 17</td>
</tr>
<tr>
<td>Expert Consensus on the Identification, Diagnosis, and Treatment of Neurotrophic Keratopathy</td>
<td>Virtual</td>
<td>11 (4 female, 7 male)</td>
<td>7 hours</td>
<td>735</td>
<td>40, 6</td>
</tr>
<tr>
<td>Estimates of Stage-Specific Preclinical Sojourn Time across 21 Cancer Types</td>
<td>Virtual</td>
<td>10 (4 female, 6 male)</td>
<td>6 hours</td>
<td>624</td>
<td>13, 1</td>
</tr>
<tr>
<td>Tapering Thrombopoietin Receptor Agonists in Primary Immune Thrombocytopenia: Expert Consensus Based on the RAND/UCLA Modified Delphi Panel Method</td>
<td>Virtual</td>
<td>10 (4 female, 6 male)</td>
<td>6 hours</td>
<td>432</td>
<td>20, 10</td>
</tr>
<tr>
<td>Developing an Emergency Department Order Set for Sickle Cell Disease in Acute Pain</td>
<td>In person</td>
<td>10 (9 female, 1 male)</td>
<td>1–2 days</td>
<td>606</td>
<td>67, 43</td>
</tr>
<tr>
<td>Development of a Severity Classification System for Sickle Cell Disease</td>
<td>In person</td>
<td>10 (6 female, 4 male)</td>
<td>1–2 days</td>
<td>640</td>
<td>64, 32</td>
</tr>
<tr>
<td>Guidelines for Management of Urgent Symptoms in Patients with Cholangiocarcinoma and Biliary Stents or Catheters Using the Modified RAND/UCLA Delphi Process</td>
<td>In person</td>
<td>15 (3 female, 12 male)</td>
<td>1 day (6 hours)</td>
<td>480 (first-round), 288 (second-round)</td>
<td>38, 10</td>
</tr>
<tr>
<td>Follow-up Intervals in Patients with Cushing's Disease: Recommendations from a Panel of Experienced Pituitary Clinicians</td>
<td>In person</td>
<td>11 (6 female, 5 male)</td>
<td>1 day (6 hours)</td>
<td>85 (first-round), 79 (second-round)</td>
<td>34, 11</td>
</tr>
</tbody>
</table>

**Notes:** A change in number of items represent expert-suggested alterations to the survey after the panel discussion. Higher numbers are percent disagreement in first-round ratings; lower numbers are percent disagreement in second-round ratings; Time spent logged into the virtual meeting. Panelists typically travel to the meeting location the evening before the meeting and spend a full day (with breaks) in person at the meeting.
disagreement following the discussion) as well as additional advantages such as shorter meeting times scheduled around clinician schedules and the ability to include diverse, sometimes international, experts who may not have otherwise been able to attend a 1–2 day in-person meeting.

Online adaptations of this method are not new. RAND developed an online elicitation system (ExpertLens), based on the original Delphi panel method, to facilitate rounds of online surveys among a large group of experts. Analytic methods are used to aggregate the data in real-time and determine what the group thinks. Like our adaptation, the authors note that this method allows for a cost-efficient way of soliciting opinions from non-collocated stakeholders.

RAND also developed the RAND/PPMD Patient-Centered Method, an online approach to engaging patients and clinicians in clinical practice guideline development. Following the Delphi panel design, the method is consistent with how clinicians and researchers typically develop practice guidelines, but also allows for the input of patients and caregivers.

Other adaptations of the method have also been published. For example, researchers combined the more traditional Delphi panel method (repeated surveys) with a follow-up in-person meeting to develop a list of standard items that should be included in health economics analysis plans (protocols detailing procedures and statistical analyses to be conducted in health economics research). The researchers used a larger group to complete two rounds of surveys (62 participants in a first-round survey and 48 in a second-round survey) and convened a smaller group of nine experts to discuss the items in-person.

Unlike these methods, our adaptation maintains the steps of the original Delphi panel process, but rather than meeting in person, the meeting is held virtually. We believe this virtual adaptation is a suitable alternative to holding meetings in person.

Conclusion
The RAND/UCLA modified Delphi panel method remains the best studied and most methodologically rigorous way to help health experts reach consensus on complex, clinical topics. The method reduces cognitive biases and efficiently harnesses the collective knowledge of experts. While other variations of this method have been asynchronous, our virtual adaptation maintains the panel meeting, allowing for valuable real-time discussion that is integral to the RAND/UCLA modified Delphi panel methodology. Moving from in-person to virtual meetings was not without challenges, but there have also been unexpected advantages. Even as COVID-19 cases decline and it will be safe to meet in person again, virtual meetings will persist. The methodology we describe here can be a cost-effective and efficient alternative for researchers and clinicians. We anticipate conducting future panels using both methods depending on the circumstances.

Author Contributions
All authors (MSB, SNG, IY) contributed significantly to this work, ensured the integrity of the work, and meet the following ICMJE criteria for authorship:

1. Made a significant contribution to the work reported, including conception, study design, execution, acquisition of data, analysis and interpretation.
2. Drafted or substantially revised or critically reviewed the article.
3. Agreed on the journal to which the article will be submitted.
4. Reviewed and agreed on all versions of the article before submission and during revision.
5. Agree to take responsibility and be accountable for the contents of the article.

MSB, SNG, and IY designed the study; acquired, analyzed, and interpreted the data; and drafted the manuscript. All authors provided final approval for the manuscript to be published and agree to be accountable for all aspects of the work.

Funding
No funding was provided for this study.
Disclosure
All authors (MSB, SNG, IY) are employees of Partnership for Health Analytic Research (PHAR), LLC, a health services research consulting firm. The authors report no other conflicts of interest in this work.

References
5. Herman PM, Hurwitz EL, Shekelle PG, Whitley MD, Couter ID. Clinical scenarios for which spinal mobilization and manipulation are considered by an expert panel to be inappropriate (and appropriate) for patients with chronic low back pain. *Med Care*. 2019;57(5):391–398. doi:10.1097/MRL.000000000001018


