

**Acceptability of an Online Modified-Delphi Panel Approach  
 for Developing Health Services Performance Measures:  
 Results from Three Panels on Arthritis Research**

Journal:	<i>Journal of Evaluation in Clinical Practice</i>
Manuscript ID	JECP-2016-0117.R1
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Khodyakov, Dmitry; RAND Corporation Grant, Sean Barber, Claire Marshall, Deborah; University of Calgary, Department of Community Health Sciences Esdaile, John Lacaille, Diane
Keywords:	health services research, healthcare
Abstract:	<p>Rationale, aims, and objectives: Online modified-Delphi (OMD) panel approaches can be used to engage large and diverse groups of clinical experts and stakeholders in developing health services performance measures. Such approaches are increasing in popularity among health researchers. However, information about their acceptability to participating experts and stakeholders is lacking but important to determine before recommending widespread use of online approaches. Therefore, the objective of this paper is to explore acceptability of the OMD panel approach from the participants' perspective.</p> <p>Method: We use data from participants in three OMD panels designed to develop performance measures for use in arthritis research and quality improvement efforts. At the end of each online panel, we surveyed clinical experts and stakeholders who shared their experiences with the OMD process by answering 13 close-ended questions using 7-point Likert-type scales. A mean of 5 or higher on a given question was treated as an indication of acceptability.</p> <p>Results: Ninety-eight clinical experts and stakeholders (92% participation rate) answered survey questions about the online process. They considered the OMD panel approach to be acceptable, particularly the ease of using the online system (mean=5.3, standard deviation=1.3) and the understanding gained from online discussions (mean=5.2, standard deviation=1.0). Participants also felt that participation in the Delphi study was interesting (mean=5.6, standard deviation=1.1).</p> <p>Conclusion(s): These findings illustrate likely acceptability and a potential for a more widespread use of OMD panel approaches by stakeholders in developing health services performance measures.</p>

--

SCHOLARONE™  
Manuscripts

For peer review only

**Acceptability of an Online Modified-Delphi Panel Approach  
for Developing Health Services Performance Measures: Results from Three Panels on  
Arthritis Research**

**Abstract**

**Rationale, aims, and objectives:** Online modified-Delphi (OMD) panel approaches can be used to engage large and diverse groups of clinical experts and stakeholders in developing health services performance measures. Such approaches are increasing in popularity among health researchers. However, information about their acceptability to participating experts and stakeholders is lacking but important to determine before recommending widespread use of online approaches. Therefore, the objective of this paper is to explore acceptability of the OMD panel approach from the participants' perspective.

**Method:** We use data from participants in three OMD panels designed to develop performance measures for use in arthritis research and quality improvement efforts. At the end of each online panel, we surveyed clinical experts and stakeholders who shared their experiences with the OMD process by answering 13 close-ended questions using 7-point Likert-type scales. A mean of 5 or higher on a given question was treated as an indication of acceptability.

**Results:** Ninety-eight clinical experts and stakeholders (92% participation rate) answered survey questions about the online process. They considered the OMD panel approach to be acceptable, particularly the ease of using the online system (mean=5.3, standard deviation=1.3) and the understanding gained from online discussions (mean=5.2, standard deviation=1.0). Participants also felt that participation in the Delphi study was interesting (mean=5.6, standard deviation=1.1).

**Conclusion(s):** These findings illustrate likely acceptability and a potential for a more

widespread use of OMD panel approaches by stakeholders in developing health services performance measures.

For peer review only

## INTRODUCTION

Developing health services performance measures is a crucial step in defining standards of care and determining its quality [1]. This process should incorporate evidence from high quality research and clinical practice guidelines, as well as be informed by the views of diverse stakeholders, including health services experts, clinicians, patients, and caregivers [2]. Delphi-based processes have long been used to select criteria that define care quality [3-6]. The Delphi method complements the results of systematic evidence reviews with consensus-focused engagement of experts and stakeholders in emerging areas where there is a lack of rigorous research, or where consensus is needed on how to apply research findings in health systems [7, 8].

Delphi processes used to develop performance measures are often called modified-Delphi because they include discussions between two or more rating rounds [4]. After reviewing the available evidence, participants rate proposed measures on different criteria (e.g., validity, feasibility), receive statistical feedback on how their responses compared to those of other participants, participate in a moderated in-person discussion of results, and revise their original answers in light of feedback and discussion [9]. Administering modified-Delphi panels, however, is logistically difficult, time-consuming, and expensive because of the need to coordinate participants' schedules, organize travel, collate responses, and manually generate individualized participant reports.

Online panels provide a possible alternative to in-person Delphi processes. Compared to in-person panels, online modified-Delphi (OMD) panels tend to offer several advantages [10, 11]. First, because participants are not required to travel to a centralized location, they provide an effective and cost-efficient format for engaging large, diverse, and geographically distributed

groups of individuals with relevant expertise [12]. Second, they automatically analyze data and generate individualized participant reports, which can significantly expedite consensus development and reduce data collection costs [13]. Third, they use unique, anonymous identifiers for participants during the online discussion to facilitate a freer exchange of ideas in groups that include clinicians and patients; anonymity can mitigate dominance of the group by a small number of vocal participants—a limitation of in-person discussions [14].

Potential disadvantages of OMD panels, however, may include lower perceived participant engagement, discussion quality, and interactivity, as well as perceived difficulty in using new technology [11, 15]. These disadvantages make acceptability [16] of OMD panels, especially their discussion components, an important indicator of their effectiveness.

Acceptability of the OMD panel approach can be operationalized as participants' experiences with the Delphi study itself (e.g., its topic, length, level of effort required, etc), online discussions, and the OMD system used for data collection purposes. Higher acceptability may lead to increased interest in the new methodology among stakeholders, increased participant motivation to engage with other panelists and answer all questions, and ultimately – increased willingness to participate in future OMD panels [17]. Therefore, it is important to determine OMD panel acceptability before promoting online processes as an option for developing performance measures and as a viable supplement to in-person panels.

In this manuscript, we examine the acceptability of an online modified-Delphi approach using data from three online panels on arthritis, explore whether perceived acceptability varied by study and by stakeholder background (i.e., physician vs. non-physician), and offer lessons learned about using this method for performance measure development.

## METHODS

We combined individual responses to the same set of survey questions about participant experiences with the online process from three OMD panels on arthritis care (see Box 1) that were conducted using the same online platform. Study 1 was an international panel tasked to develop cardiovascular quality indicators for rheumatoid arthritis [18]. Study 2 was a national panel to develop system-level performance measures for evaluating models of care for inflammatory arthritis in Canada [19]. Study 3 was a provincial panel designed to develop key performance indicators for evaluating centralized intake for arthritis care in Alberta, Canada [20]. All studies were approved by the RAND Human Subjects Protection Committee.

### **Study Participants**

Stakeholders with relevant expertise identified from professional organizations were recruited via email and consented to participation. Based on previous research on online expert panels, the intended sample size for each panel was 20-40 participants [13]. To reach this target, a diverse group of 43 stakeholders from North America and Europe was invited to participate in Study 1. In Study 2, 50 arthritis stakeholders from across Canada were invited. In Study 3, 28 stakeholders from Alberta were invited. All panels involved physicians, allied health professionals, researchers, and patients; Study 3 also included clinic managers/administrators.

### **Research Design**

All panels were conducted using RAND's ExpertLens™ - an OMD platform that combines rounds of questions with a round of statistical feedback and online discussion and automatically analyzes group responses [14]. Although ExpertLens has been used successfully in numerous studies on different healthcare topics [21-27], these three panels were the first studies

to use ExpertLens for developing health services performance measures. Each OMD round occurred over 7-14 days. Round deadlines were extended based on participation rates. Participants received periodic reminders to maximize engagement. No financial incentives were offered to participants; expenses associated with attending an in-person meeting in Study 3 were reimbursed.

In Round 1, participants rated measures on a set of four to six Likert scales and explained their responses using open text boxes. In Round 2, medians and quartiles were calculated for each Round 1 question and displayed on histograms that showed the frequency of all participants' responses in addition to a participant's own response to that question. Participants then engaged in an asynchronous and anonymous discussion of Round 1 results using an online forum. CEHB, a rheumatologist and health services researcher, moderated online discussions in all three panels. By design, there was also a face-to-face meeting in addition to the online forum in Study 3. In Round 3, participants revised their Round 1 responses based on Round 2 feedback and discussion. They also completed an optional survey about their experiences with the online process [13]. Substantive panel findings have been published separately [18-20].

### **Measures and Analysis**

At the end of Round 3, participants used 7-point Likert scales (1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Neutral, 5= Slightly Agree, 6=Agree, 7=Strongly Agree) to rate statements describing their experiences with the Delphi study, online discussions, and the online system, which we treated as indicators of acceptability of the OMD panel approach (see Table 1). Although not formally validated, these statements were based on research on computer-mediated communication and factors that may affect participant experiences in online panels



[28-30] and used in an earlier project [13]. For each study and the combined dataset, we calculated means and standard deviations for each acceptability item. Missing values (1.4% of items in Study 1; 0.5% of items in Study 2; 0.1% of items in Study 3) were imputed with a value of 4 (neutral response category). As in previous research, we used the mean values and original labels on the 7-point scale to describe participants' experiences [13]; a mean  $\geq 5$  on positively worded items (and  $\leq 3$  on negatively worded items) indicated an "acceptable" response. To facilitate interpretation, we rounded mean values. To ensure robustness of our findings, we calculated the percent of participants with negative (ratings 1-3), neutral (rating of 4), and positive (ratings of 5-7) opinions about the acceptability of the online process across the three studies. For the latter analyses, we reverse coded negatively worded items (i.e., "1" corresponded to the most negative and "7" to the most positive participant experience). If the majority (more than 51%) of responses fell into a positive category, we considered this to be a sign of agreement on acceptability of OMD panels [31, 32]. Finally, we also conducted ANOVA analyses to examine whether average responses for each item were different by study and by stakeholder background (i.e., physician vs. non-physician) given the influence of professional background on panel results [33].

## RESULTS

Out of 121 invitees across the three panels, 107 experts (88%) participated in one of the OMD panels (see Table 1). Of 107 participants, 76 (71%) accessed Round 2, with 37 (49%) individuals making 212 online comments. In Study 3, 12 participants (44%) contributed to the online discussion and 19 (70%) attended an in-person meeting during Round 2. Of 107 participating experts, 98 (92%) answered acceptability questions. While the majority of

participants in the first two studies were physicians, only a third of Study 3 participants were physicians and roughly a quarter were researchers (see Table 2).

Participants considered the OMD panel approach to be acceptable (see Table 3, All Studies column). Participants agreed that their participation in the study was interesting ( $M=5.6$ ,  $SD=1.1$ ), and they did not find it frustrating ( $M=2.8$ ,  $SD=1.3$ ). They had a neutral opinion about the study length ( $M=3.7$ ,  $SD=1.5$ ) and the effort needed to complete it ( $M=3.8$ ,  $SD=1.5$ ). Regarding the discussion round, participants slightly agreed that the discussion caused them to revise their original answers ( $M=4.7$ ,  $SD=1.3$ ), brought out views they hadn't considered ( $M=4.8$ ,  $SD=1.2$ ), brought out divergent views ( $M=4.9$ ,  $SD=1.0$ ), and gave them a better understanding of the issues discussed ( $M=5.2$ ,  $SD=1.0$ ). They also slightly agreed that participants debated each others' viewpoints during the discussions ( $M=4.6$ ,  $SD=1.3$ ), and that they were comfortable expressing their views during the discussion round ( $M=5.2$ ,  $SD=1.3$ ). Participants did not indicate difficulty following discussions ( $M=3.5$ ,  $SD=1.5$ ). Lastly, participants slightly agreed that the online system itself was easy to use ( $M=5.3$ ,  $SD=1.3$ ), and they would like to use it in the future ( $M=4.9$ ,  $SD=1.3$ ). Results presented in Table 4 support these findings: more than 51% of participants chose a response category that fell within the range of acceptable responses on 10 out of 13 items.

When comparing acceptability between studies, there were significant differences on four items (see Table 3). First, compared to Study 1 participants, Study 2 and Study 3 participants were more likely to agree that the study was too long ( $p=0.018$ ). Second, compared to Study 3 participants, participants in Study 1 and Study 2 were less likely to agree that their participation took a lot of effort ( $p=0.004$ ). Third, participants in Study 2 were more likely to agree that the discussion rounds caused them to revise their original answers, compared to Study 1 participants

( $p=0.005$ ). Finally, compared to Study 1 participants, Study 3 participants were more likely to agree that the discussions brought out divergent views ( $p=0.014$ ).

Only three items differed when comparing acceptability by stakeholder background, and all three items related to Round 2 discussions. Specifically, non-physicians were statistically significantly more likely than physicians to agree that participants debated each others' viewpoints during the discussions ( $p=0.037$ ), the discussions brought out divergent views ( $p=0.046$ ), and the discussion round caused them to revise their original answers ( $p=0.029$ ) (data not shown in a table).

## DISCUSSION

OMD approaches have the potential to facilitate engagement of large and diverse stakeholder groups in developing performance measures. We explored acceptability of this approach using data from three panels that developed performance measures aimed to inform arthritis care. Participants found the online approach to be acceptable, particularly the ease of the online system's use and the understanding gained from online discussions. In line with previous studies investigating satisfaction with OMD processes [13, 15], these findings contribute to growing research on the potential for more widespread use of the online approach for systematic development of performance measures that reflect important stakeholder perspectives [34]. Because participants do not have to travel to an in-person meeting and organizers do not manually generate individualized reports to each participant, reduced study costs allow for efficient engagement of a broader range and a greater number of stakeholders within a short period of time. Study 1 in particular would not have been possible due to high international travel costs.

Although deemed generally acceptable in all three studies, there were some differences in acceptability ratings of the OMD approach between the studies. For example, Study 1 participants disagreed that their study was too long, whereas Study 2 and Study 3 participants had a neutral opinion about their study length. Although Study 2 had the smallest number of total questions asked, participants had to rate items on six different criteria, which might have increased a feeling of participation burden. Study 3 was the longest and most complex study because participants had to participate in online and in-person discussions and rate the largest number of measures. Study 3 participants, however, were more likely than Study 1 participants to agree that discussions brought out divergent views—possibly because they took part in online and in-person discussions. Moreover, Study 2 participants were more likely to agree that the discussion round helped them revise their original responses, which could be explained by more complex measures in Study 2 and thus greater utility of others' feedback on participants' initial Round 1 responses. Lastly, physicians were less likely to agree with items related to Round 2 discussions leading to revised ratings in Round 3, which corresponds to the theory that strongly held views by those with trained specialist knowledge are more difficult to change via consensus-based processes [33].

In interpreting these findings based on our personal experiences administrating these panels, we offer three lessons for those interested in using OMD approaches for developing performance measures. First, it is important to find the right number of performance measures and rating criteria to ensure that participants do not find the online process to be too burdensome. Our experience shows that approximately 10-12 performance measures and five rating criteria may provide a manageable number of questions to answer and still generate enough useful data. Second, it is important to aim to keep OMD processes brief and efficient to ensure that

participants do not perceive them to be overly lengthy and burdensome. Study organizers should provide sufficient time to participants to complete each round; extending round deadlines based on current participation rates was needed in all three studies. Finally, it must be noted that, while providing a more efficient process for developing performance measures than in-person meetings, online discussions may be supplemented with in-person discussions to more effectively engage panelists in discussing complex issues. Participants in Study 3, a small local study that had both online and in-person discussions, were more likely to report that discussions brought out divergent views, especially compared to Study 1 participants. Therefore, organizers of such panels may want to explore the feasibility of combining online and in-person/phone discussions, while conducting rating rounds completely online. This hybrid format makes it easier to analyze rating data and discuss complex issues in depth.

Several limitations should be considered when interpreting these results. First, our results may have limited generalizability because acceptability data came from studies conducted by the same team of investigators that focused on arthritis. We note, however, that results from all three panels, which varied in length, complexity, scope (e.g., regional, national, and international), and topic support the acceptability of the online approach. Moreover, while participant experiences may depend on the expertise of the team of investigators conducting the study, the lessons learned described above can help researchers ensure positive participant experiences. Second, in selecting panelists, we purposefully attempted to include diverse stakeholders with relevant expertise but may not have been representative of all stakeholders. Moreover, 12% of invited panelists did not participate in any study. 8% of participating panelists did not answer acceptability questions, possibly because they might have had poorer study experiences. Third, very limited demographic information was collected in all studies to ensure participant

anonymity to the research team. Future studies should explore differences in acceptability ratings by participants' demographic characteristics. Fourth, there are no commonly used measures of acceptability of online Delphi processes. Although we used previously published measures developed based on the literature on computer-mediated communication [13], additional research should explore the dimensions underlying acceptability of online panels using factor analysis. Finally, we do not have similar data from in-person panels on this topic. Future research should seek to compare acceptability of online and in-person panels using the same instrument.

In summary, this study illustrates the acceptability of the OMD panel approach for engaging healthcare stakeholders in developing performance measures. The use of acceptable online approach can allow a large number of diverse participants from a wide geographical distribution to be assembled, as well as equal participation of stakeholders without dominance of discussion by those with more perceived authority, due to anonymity of the discussion. Those involved in developing measures should consider online approaches to ensure the input of all relevant stakeholders is reflected.

## ACKNOWLEDGMENTS

The authors wish to acknowledge Nathaly Pacheco-Santivanez for assistance with panel administration.

## FUNDING

The studies discussed in this manuscript were funded by a Canadian Institutes for Health Research (CIHR) grant (129628), a CIHR Planning Grant 2013-10-15 (Funding reference number is 218913), a CIHR Planning Grant (20132PLH, funded through Priority Announcement Health Services and Policy Research), a Partnership for Research and Innovation in the Health System (PRIHS) grant (201300472) from Alberta Innovates – Health Solutions (AIHS) and Alberta Health Services (AHS), and funding and in-kind resources provided by the Arthritis Alliance of Canada (AAC). Additional funding was provided by RAND Corporation.

## REFERENCES

1. McGlynn EA, Asch SM. Developing a clinical performance measure. *American journal of preventive medicine*. 1998;14(3 Suppl):14-21.
2. Kotter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators: A systematic review. *Implementation Science*. 2012;7:21.
3. Kötter T, Schaefer FA, Scherer M, Blozik E. Involving patients in quality indicator development: A systematic review. *Patient Preference and Adherence*. 2013;7:259.
4. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS one*. 2011;6(6):e20476.
5. Normand S-LT, McNeil BJ, Peterson LE, Palmer RH. Eliciting expert opinion using the Delphi technique: identifying performance indicators for cardiovascular disease. *International Journal for Quality in Health Care*. 1998;10(3):247-60.
6. Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J, Marres H, Dijkstra R, Grol R. Clinical indicators: development and applications. *Neth J Med*. 2007;65(1):15-22.
7. Scott DA, Reeves C, Bate A, Van Teijlingen ER, Russell EM, Napper M, Robb CM, (2001). Eliciting public preferences for healthcare: A systematic review of techniques. *Health Technology Assessment*. 2001;5(5):1-186.
8. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8(1):e1000393.



9. Okoli C, Pawlowski S. The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*. 2004;42(1):15-29.
10. Snyder-Halpern R, Thompson C, Schaffer J, editors. Comparison of mailed vs. Internet applications of the Delphi technique in clinical informatics research. *Proceedings of the AMIA Symposium*; 2000. American Medical Informatics Association.
11. Donohoe H, Stellefson M, Tennant B. Advantages and limitations of the e-Delphi technique: Implications for health education researchers. *American Journal of Health Education*. 2012;43(1):38-46.
12. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, Thomson R, Barratt A, Barry M, Bernstein S. Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *BMJ*. 2006;333(7565):417-23.
13. Khodyakov D, Hempel S, Rubenstein L, Shekelle P, Foy R, Salem-Schatz S, O'Neill S, Danz M, Dalal S. Conducting Online Expert panels: A feasibility and experimental replicability study. *BMC Medical Research Methodology*. 2011;11(1):174.
14. Dalal S, Khodyakov D, Srinivasan R, Straus S, Adams J. ExpertLens: A system for eliciting opinions from a large pool of non-located experts with diverse knowledge. *Technol Forecast Soc*. 2011;78(8):1426-44.
15. Deshpande AM, Shiffman RN, Nadkarni PM. Metadata-driven Delphi rating on the Internet. *Computer Methods and Programs in Biomedicine*. 2005;77(1):49-56.
16. Ayala GX, Elder JP. Qualitative methods to ensure acceptability of behavioral and social interventions to the target population. *Journal of Public Health Dentistry*. 2011;71:S69-S79.
17. Bolger F, Wright G. Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting & Social Change*. 2011;78:1500-13.

18. Barber C, Marshall D, Alvarez N, Mancini GB, Lacaille D, Keeling S, Avina-Zubieta JA, Khodyakov D, Barnabe C, Faris P, Smith A, Noormohamed R, Hazlewood G, Martin L, Esdaile J. Development of Cardiovascular Quality Indicators for Rheumatoid Arthritis: Results from an International Expert Panel Using a Novel Online Process. *The Journal of Rheumatology*. 2015;42(9):1548-55.
19. Barber C, Marshall D, Mosher D, Akhavan P, Tucker L, Houghton K, Batthish M, Levy D, Schmeling H, Ellsworth J, Tibollo H, Grant S, Khodyakov D, Lacaille D. Development of System-Level Performance Measures for Evaluation of Models of Care for Inflammatory Arthritis in Canada. *The Journal of Rheumatology*. 2016;doi: 10.3899/jrheum.150839.
20. Barber C, Patel JN, Woodhouse L, Smith C, Weiss S, Homik J, LeClerc S, Mosher DP, Christiansen T, Howden JS, Wasylak T, Greenwood-Le JM, Emrick A, Suter E, Kathol B, Khodyakov D, Grant S, Campbell-Scherer D, Phillips L, Hendricks J, Marshall D. Development of Key Performance Indicators to Evaluate Centralized Intake for Patients with Osteoarthritis and Rheumatoid Arthritis. *Arthritis Research & Therapy*. 2015;17(322):1-12.
21. Ohno-Machado L, Agha Z, Bell DS, Dahm L, Day ME, Doctor JN, Gabriel D, Frey LJ. pSCANNER: patient-centered Scalable National Network for Effectiveness Research. *Journal of the American Medical Informatics Association*. 2014;21(4):621-6.
22. Jones MM, Pickett J, Chataway J, Swartz J, Yaqub O, Smith P, Palar K, Terlikowski J, Mark D, McColl W, Hackett P, Manville C, Glick P. Mapping pathways: Developing evidence-based, people-centred strategies for the use of antiretrovirals as prevention. Cambridge, UK: RAND Europe; 2013.
23. Khodyakov D, Savitsky TD, Dalal S. Collaborative learning framework for online stakeholder engagement. *Health Expectations*. 2015;DOI: 10.1111/hex.12383.

24. Rubenstein L, Khodyakov D, Hempel S, Danz M, Salem-Schatz S, Foy R, O'Neill S, Dalal S, Shekelle P. How can we recognize continuous quality improvement? *International Journal for Quality in Health Care*. 2014;26(1):6-15.
25. Claassen CA, Pearson JL, Khodyakov D, Satow PM, Gebbia R, Berman AL, Reidenberg DJ, Feldman S, Molock S, Carras M, Lento R, Sherrill J, Pringle B, Dalal S, Insel TR. Reducing the burden of suicide in the U.S.: The aspirational research goals of the National Action Alliance for Suicide Prevention Research Prioritization Task Force. *American Journal of Preventive Medicine*. 2014;47(3):309-14.
26. Khodyakov D, Mikesell L, Schraiber R, Booth M, Bromley E. On using ethical principles of community-engaged research in translational science. *Translational Research*. 2016;DOI: <http://dx.doi.org/10.1016/j.trsl.2015.12.008>.
27. Khodyakov D, Stockdale S, Smith N, Booth M, Altman L, Rubenstein L. Patient engagement in the process of planning and designing outpatient care improvements at the Veterans Administration Healthcare System: Findings from an online expert panel. *Health Expectations*. In Press.
28. Olaniran BA. A model of group satisfaction in computer-mediated communication and face-to-face meetings. *Behaviour & Information Technology*. 1996;15(1):24-36.
29. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*. 1983;29(5):530-45.
30. Hiltz SR, Johnson K. User satisfaction with computer-mediated communication systems. *Management Science*. 1990;36(6):739-64.
31. McKenna HP. The selection by ward managers of an appropriate nursing model for long-stay psychiatric patient care. *Journal of Advanced Nursing*. 1989;14(9):762-75.

32. Loughlin KG, Moore LF. Using Delphi to achieve congruent objectives and activities in a pediatrics department. *Academic Medicine*. 1979;54(2):101-6.
33. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2(3):1-88.
34. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *Quality and Safety in Health Care*. 2002;11(4):358-64.

**Box 1. Study Characteristics**

<b>Study Information</b>	<b>Study 1</b>	<b>Study 2</b>	<b>Study 3</b>
Study Dates	11/04/13-12/03/13	09/22/14-11/06/14	01/12/15-02/11/15
<b><i>Number of Items and Rating Criteria</i></b>			
Number of items (quality indicators) rated	11	6	31
Number of rating criteria	4	6	5
<b><i>Rating Criteria Used*</i></b>			
Validity: Overall	X	X	X
Validity: Reflects Quality Health System			X
Validity: Importance			X
Feasibility: Overall	X		
Feasibility: Ability to Control		X	
Feasibility: Availability of Required Information		X	X
Feasibility: Reliability		X	
Relevance	X	X	
Likelihood of Use	X	X	X

\*Participants rated proposed performance measures on various criteria related to their validity, feasibility, relevance, and likelihood of use. Each “X” above indicates that a given study has used the rating criteria listed in the first column. Full descriptions of all rating criteria can be found in Online Supplement 1.

Table 1: Study Participation Rates

Study Stage	All Studies	Study 1	Study 2	Study 3
1) Number of potential panelists we invited to participate	121	43	50	28
2) Number of invited panelists who participated (% of potential panelists we invited)	107 (88%)	37 (86%)	43 (86%)	27 (96%)
3) Number of participants who accessed the Round 2 discussion (% of number of participants)	76 (71%)	32 (87%)	32 (74%)	12 (44%) online 19 (70%) in-person
4) Number of participants who posted a comment in Round 2 (% of number of participants who accessed Round 2)	37 (49%)	19 (59%)	14 (44%)	4 (33%)
5) Number of comments in Round 2 (average per participant who posted a comment in Round 2)	212 (6)	110 (6)	72 (5)	30 (8)
6) Number of participants who completed post-panel acceptability questionnaire (% of number of participants)	98 (92%)	32 (74%)	43 (100%)	23 (79%)

This table discusses the flow of participation within each study: 1) recruitment, 2) enrollment in Round 1, 3) accessing Round 2 (i.e., logging into ExpertLens in Round 2), 4 and 5) posting a comment in Round 2, and 6) completing the acceptability questionnaire after Round 3.

**Table 2. Participant Demographics**

Variable	All Studies (n = 98)	Study 1 (n = 32)	Study 2 (n = 43)	Study 3 (n = 23)
<i>Professional Background n (%)</i>				
Physician	56 (57%)	26 (81%)	22 (51%)	8 (35%)
Allied health professional	14 (14%)	2 (6%)	10 (23%)	2 (9%)
Healthcare/Clinic Manager	4 (4%)	--	--	4 (17%)
Methodologist/researcher	10 (10%)	1 (3%)	4 (9%)	5 (22%)
Patient/consumer/advocate	8 (8%)	2 (6%)	3 (7%)	3 (13%)
Other	4 (4%)	1 (3%)	2 (5%)	1 (4%)
Did not report	2 (2%)	--	2 (5%)	--

This table provides data on the self-reported professional background of participants who completed the post-panel acceptability questionnaire

**Table 3. Online Modified-Delphi Acceptability Ratings**

Item	Mean (SD)			
	All Studies	Study 1	Study 2	Study 3
<b><i>Delphi Study Acceptability – Positively Worded Item</i></b>				
1. Participation in this study was interesting	5.6 (1.1)	5.7 (1.4)	5.5 (0.9)	5.4 (1.1)
<b><i>Delphi Study Acceptability – Negatively Worded Items</i></b>				
2. This study was too long	3.7 (1.5)*	3.1 (1.5) <sup>a,b</sup>	4.1 (1.4) <sup>a</sup>	3.9 (1.4) <sup>b</sup>
3. Participation in this study was frustrating	2.8 (1.3)	2.5 (1.3)	2.9 (1.2)	3.0 (1.5)
4. Participation in this study took a lot of effort	3.8 (1.5)**	3.2 (1.34) <sup>b</sup>	3.7 (1.4) <sup>c</sup>	4.6 (1.5) <sup>b,c</sup>
<b><i>Round 2 Online Discussion Acceptability – Positively Worded Items</i></b>				
5. The discussions gave me a better understanding of the issues	5.2 (1.0)	5.2 (1.0)	5.1 (1.1)	5.4 (1.1)
6. Participants debated each others' viewpoints during the discussions	4.6 (1.3)	4.3 (1.3)	4.8 (1.1)	4.8 (1.4)
7. The discussions brought out views I hadn't considered	4.8 (1.2)	4.5 (1.4)	4.9 (1.1)	4.9 (1.1)
8. The discussions brought out divergent views	4.9 (1.0)*	4.6 (1.0) <sup>b</sup>	5.0 (0.9)	5.4 (0.9) <sup>b</sup>
9. The discussion round caused me to revise my original answers	4.7 (1.3)**	4.1 (1.6) <sup>a</sup>	5.1 (0.9) <sup>a</sup>	4.7 (1.1)
10. I was comfortable expressing my views in the discussion round	5.2 (1.3)	5.3 (1.4)	5.0 (1.3)	5.3 (1.1)
<b><i>Round 2 Online Discussion Acceptability – Negatively Worded Item</i></b>				
11. I had trouble following the discussions	3.5 (1.5)	3.9 (1.6)	3.6 (1.3)	3.0 (1.3)
<b><i>Online System Acceptability – Positively Worded Items</i></b>				
12. The ExpertLens system was easy to use	5.3 (1.3)	5.2 (1.3)	5.5 (1.3)	5.2 (1.2)
13. I would like to use ExpertLens in the future	4.9 (1.3)	4.9 (1.6)	4.9 (1.2)	4.9 (1.3)

This table reports the mean (SD) rating participants gave on the acceptability of the online modified-Delphi panel approach for developing health services performance measure. Participants used a 7-point Likert scale: 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Neutral, 5=Slightly Agree, 6=Agree, and 7=Strongly Agree. For positively worded items, higher scores indicate greater acceptability. For negatively worded items, lower scores indicate greater acceptability.

\*indicates  $p \leq 0.05$  and \*\* indicates  $p \leq 0.01$  for ANOVA analyses examining whether average responses for each item differed by study: <sup>a</sup> indicates scores for item differed between Study 1 and Study 2, <sup>b</sup> between Study 1 and Study 3, and <sup>c</sup> between Study 2 and Study 3.



**Table 4. Acceptability Questionnaire Scores (All Studies)**

Item	Percent Rating		
	1 - 3 (Unacceptable)	4 (Neutral)	5-7 (Acceptable)
<b><i>Delphi Study Acceptability</i></b>			
1. This study was too long*	31 (35.3%)	22 (25.0%)	35 (39.8%)
2. Participation in this study was frustrating*	9 (10.2%)	20 (22.7%)	59 (67.1%)
3. Participation in this study took a lot of effort*	30 (34.1%)	23 (26.1%)	35 (39.7%)
4. Participation in this study was interesting	2 (2.2%)	13 (14.8%)	73 (83.0%)
<b><i>Round 2 Online Discussion Acceptability</i></b>			
5. I had trouble following the discussions*	27 (30.7%)	19 (21.6%)	42 (47.7%)
6. The discussions gave me a better understanding of the issues	3 (3.4%)	20 (22.7%)	65 (73.9%)
7. Participants debated each others' viewpoints during the discussions	15 (17.1%)	25 (28.4%)	48 (54.5%)
8. The discussions brought out views I hadn't considered	11 (12.5%)	24 (27.3%)	53 (60.2%)
9. The discussions brought out divergent views	9 (10.2%)	18 (20.5%)	61 (69.3%)
10. The discussion round caused me to revise my original answers	15 (17.1%)	13 (14.8%)	60 (68.2%)
11. I was comfortable expressing my views in the discussion round	8 (9.1%)	17 (19.3%)	63 (71.6%)
<b><i>Online System Acceptability</i></b>			
12. The ExpertLens system was easy to use	10 (11.4%)	13 (14.8%)	65 (73.8%)
13. I would like to use ExpertLens in the future	9 (10.2%)	23 (26.1%)	56 (63.6%)

This table reports the number of participants (%) who gave a particular rating for each item on the acceptability of the online modified-Delphi panel approach for developing health services performance measure. Participants used a 7-point Likert scale, where 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Neutral, 5=Slightly Agree, 6=Agree, and 7=Strongly Agree.

\*For these analyses, negatively worded items were reverse coded (i.e., "1" corresponded to the most negative and "7" to the most positive participant experience) to improve interpretability across items. We calculated the number and percent of participants with negative (ratings 1-3), neutral (rating of 4), and positive (ratings of 5-7) opinions about the acceptability of the online process across the three studies.