ROMPER: The RAND/USC OPTIC Method for Policy Expert Ratings

Sean Grant [a,b,*], Rosanna Smart [a]

[a] *RAND Corporation, 1776 Main St, Santa Monica, CA 90401, USA*
[b] *HEDCO Institute for Evidence-Based Educational Practice, College of Education, University of Oregon, 1215 University of Oregon, Eugene, OR 97403-1215, USA*

ARTICLE INFO

ABSTRACT

We developed an expert panel approach for identifying expert views on the effectiveness and implementability of population-level policy interventions. ROMPER—the RAND/USC OPTIC Method for Policy Expert Ratings—involves an online, three-round, modified-Delphi process:

- Experts rate and comment on policies according to domains of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence-to-Decision framework.
- To identify consensus on policy effectiveness and implementability, expert ratings are analyzed using the Inter-Percentile Range Adjusted for Symmetry (IPRAS) technique from the RAND/UCLA Appropriateness Method and visualized using a forest plot. To explain consensus, expert comments are analyzed using reflexive thematic analysis and reported following the Standards for Reporting Qualitative Research.
- To provide actionable information for decisionmakers, each policy is summarized in a "Policy Profile" adapted from GRADEPro Evidence-to-Decision tables.

We validated ROMPER in two studies that successfully recruited the targeted sample size, retained experts through all three rounds, and examined consensus on which policies are (not) effective and implementable. ROMPER protocols, materials, data, and code are openly available on the Open Science Framework with Creative Commons licensing for replication and reuse. ROMPER provides a validated, replicable, open access approach for eliciting expert views on both policy effectiveness and implementability—and for summarizing (lack of) consensus specifically for policymakers.

Specifications table

Subject area:	Medicine and Dentistry
More specific subject area:	Public Health and Policy Sciences
Name of your method:	ROMPER: The RAND/USC OPTIC Method for Policy Expert Ratings
Name and reference of original method:	Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science, 9*(3), 458–467.
Resource availability:	https://osf.io/gb936/

* Corresponding author.
  E-mail address: sgrant@rand.org (S. Grant).
  Social media: 🐦 @GrantSeanP (S. Grant)

Method details

Background

Ideally, recommendations and decisions about health policies are based on direct empirical evidence of effectiveness from rigorous causal research, with guidance on implementation considerations provided through rigorous implementation research [1]. In practice, guideline developers and policymakers often have insufficient evidence for instrumental use [2]; for example, in the case of policy innovations, process or impact evaluation research cannot be conducted until the policy has been tried. Consequently, rigorous evidence syntheses often need to be combined with the use of formal consensus methods to elicit expert views on health policies [3].

The Delphi technique is a formal consensus development method prominently used in medicine, health services, and allied disciplines [4]. Delphi is defined by anonymity, iteration, controlled feedback of responses to experts, and statistical aggregation of expert responses [5]. The facilitators ask an expert panel to respond anonymously to a questionnaire, aggregate their responses, provide this aggregation back to the panel, and invite panelists to revisit their responses in light of the aggregated feedback [6]. Through these procedures, Delphi aims to mitigate potential social and psychological biases among experts involved in the elicitation process [7]. Specific protocols for applying the Delphi technique have been developed for identifying expert views on the appropriateness and patient-centeredness of clinical-level interventions [8,9]. A similar protocol is needed for decision-making processes focused on policies (rather than clinical interventions). In this manuscript, we report on the development of a methodological protocol for applying the Delphi technique to identify expert views on the effectiveness and implementability of population-level policy interventions.

The rand/usc optic method for policy expert ratings

The RAND/USC OPTIC Method for Policy Expert Ratings (ROMPER) provides a validated, replicable, open access approach for eliciting expert views on policy effectiveness and implementability—and to summarize expert consensus for policymakers. ROMPER is amenable with prominent approaches for developing guidelines—i.e., Grading of Recommendations Assessment, Development and Evaluation (GRADE) and RAND/UCLA Appropriateness Method (RAM)—and can help identify potential high-value polices to adopt or low-value policies to de-implement, as well as important implementation and sub-group considerations.

We developed ROMPER for a project of the RAND-USC Opioid Policy Tools and Information Center (OPTIC) that examined variation in state-level opioid policies. ROMPER protocols, materials, data, and code are openly available with Creative Commons licensing for replication and reuse by other opioid policy researchers (https://osf.io/gb936/). We conducted an initial panel on naloxone access laws [10,11], and we replicated the method in a subsequent panel on policies relevant for linkage to and retention in medications for opioid use disorder (OUD) [12,13]. We used the Open Science Framework to prospectively register the panels on naloxone access laws (https://osf.io/e2aw5/) and OUD treatment policies (https://osf.io/3vuq6/). ROMPER methodological procedures were given exempt status by the RAND Human Subjects Protection Committee (Study 2018–0506).

In our validation studies of ROMPER, panels involved an online, three-round, modified-Delphi approach using the RAND ExpertLens online platform [14]. This approach consists of two rating rounds, with a round of online group discussion and feedback in between. Our conceptualizations of policy effectiveness and implementability are based on the GRADE Evidence-to-Decision (EtD) framework for health systems and public health, which is designed to help panels form recommendations to policymakers responsible for decision-making on behalf of a population affected by those decisions [1,15]. ROMPER involves six key steps (see Graphical Abstract). Our protocol was ruled exempt from further review by our institutional review board.

Defining policies

The first step of ROMPER involves defining the policies of interest on which experts will be queried. We recommend a taxonomic approach to defining policies that involves mutually exclusive and mutually comprehensive categories [16]. This approach involves initially organizing the policy area at the highest taxonomic level according to broad domains and then further categorizing policies at increasing levels of granularity, such as classes within a domain, families within a class, mechanisms within a family, and techniques within a mechanism (see Fig. 1). This systematic categorization of policies at multiple levels allows the research team to better describe policies, identify existing empirical evidence on the polices to provide experts as background reading, and examine whether experts believe differences in distinct policy components influence effectiveness and implementability. In our initial uses of ROMPER, we worked iteratively with our research team and advisory board to construct inclusive lists of implemented or proposed policies in a demarcated area of our taxonomy.

Defining rating criteria

Once policies are defined, criteria for rating policies are designed based on the GRADE Evidence-to-Decision Framework [1,15]. To reduce participant burden (i.e., the number of questions a participant must answer), the default approach involves conducting two panels concurrently, with each expert participating only in one panel. One set of experts rates the direction and magnitude of intervention effectiveness on outcomes related to the theory of change of the policy intervention. For example, in our panel on naloxone access laws [10], these effectiveness outcomes included the primary objective (preventing fatal opioid overdose), mechanism of action (distributing naloxone through pharmacies), and hypothesized negative consequences (increases in nonfatal overdoses and OUD prevalence). Participants provide ratings on a Likert scale [1–9], where a score of 5 means "no effect" on the outcome, scores
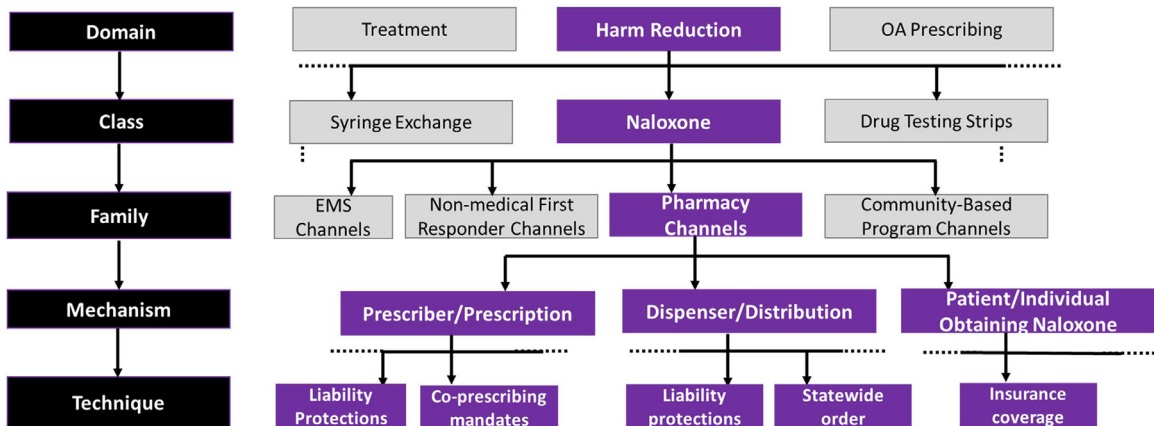
Fig. 1. Example policy taxonomy.

below 5 mean the policy reduces the outcome (lower scores indicating larger reductions), and scores above 5 mean the policy increases the outcome (higher scores indicating larger increases). Based on feedback during pilot-testing, we did not reverse the scale when a larger reduction was a negative versus positive outcome (and similarly, when a larger increase was a positive rather than negative outcome) indicates item where disagreement (lower score) signals a positive experience. Another set of experts rates the policy on explicit criteria related to key determinants of implementing policy changes [17]. For example, in our panel on opioid treatment policies, we asked participants to rate each policy on four criteria:

- Acceptability: the extent to which the policy is acceptable to the general public in the state or community where the policy has been enacted.
- Feasibility: the extent to which it is feasible for a state or community to implement the policy as intended.
- Affordability: the extent to which the resources (costs) required to implement the policy are affordable from a societal perspective.
- Equity: the extent to which the policy is equitable in its impact on health outcomes across populations of people who use opioids.

Participants provide ratings on a 9-point Likert scale, where scores of 1 to 3 mean "low," scores of 4 to 6 mean "moderate," and scores of 7 to 9 mean "high" acceptability, feasibility, affordability, or equitability.

Recruiting experts

Once policies are defined and rating criteria operationalized, the next step is selecting experts in these policies and criteria. We recommend recruiting experts using a multipronged strategy to facilitate diversity of perspectives and experience. In our initial uses of ROMPER, we first developed a list of potential participants based on published research, suggestions from our project advisory board, and member lists of relevant organizations. We then used a "snowball sampling" approach whereby stakeholders could nominate further participants. We aimed to recruit a sample with sufficient diversity of perspectives and experience: i.e., 20–40 participants per panel [18]. We approached experts from numerous stakeholder groups in the policy area: i.e., advocates, persons with lived experience, healthcare providers, human and social service practitioners, policymakers, and researchers. We sent all identified experts a recruitment survey asking them to indicate whether they would like to participate. Interested experts completed an online registration survey to provide informed consent, self-report demographic data, and indicate whether they preferred to rate policies on their effectiveness or implementability. Experts participated in their preferred panel; we assigned experts with no preference to panels in a manner that optimized balance on demographic variables. Participants received a $300 prepaid card for completing all rounds, with pro-rated honorarium for partial completion.
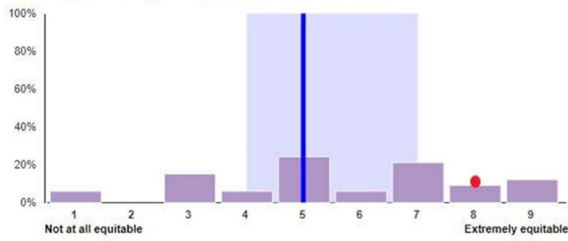
Running panels

We refined our study protocols through cognitive-testing of the questionnaire with eligible experts and pilot-testing of its programming into the ExpertLens platform with members of our research team. We concurrently conducted two online modified Delphi panels with no overlap in participants: one focused on policy effectiveness (effectiveness panel), and one focused on policy implementability (implementation panel). In Round One, participants rated either the effect of each policy on relevant outcomes (effectiveness panel) or criteria related to implementability (implementation panel). Participants could also comment on the rationale underpinning their ratings. In Round Two, experts saw graphed results from Round One (i.e., frequency histograms, median group ratings, and their own ratings). We also displayed summaries of thematic analyses of Round One comments next to each graph as well as all participant comments from Round One (see Fig. 2 for images that participants see on a single screen). Participants then explored areas of agreement

Fig. 2. Example of round two feedback and discussion.

and disagreement by discussing these results in anonymous, asynchronous online discussion boards moderated by members of our research team to ensure consistency in discussion management. The moderator can encourage participant interaction when needed; suggest new discussion topics that expand the depth and breadth of existing discussion topics; stimulate greater engagement in the topic; and encourages individuals' active discourse and critical thinking. Through this process, the participants are asked to comment on the responses of the other participants, defend their own positions, develop new ideas, or answer new questions introduced by the moderators. In Round Three, informed by Round One results and Round Two discussion, experts independently provided final ratings for each policy's effectiveness and implementability. We ensured participant anonymity of responses via use of anonymized usernames in ExpertLens. Each round took participants approximately one hour to complete and remained open for one to three weeks.

Analyzing data

Below we present information on analyzing the quantitative rating and qualitative comment data from ROMPER panels. Information on actual number of panelists completing each round can be found in the manuscripts on our ROMPER panels [10–13]. The quantitative and qualitative datasets themselves can be found on the Open Science Framework (https://osf.io/gb936/).

Quantitative Data Analysis. Our quantitative data analysis seeks to answer the question: "what is expert consensus on the effectiveness and implementability of the types of policy under evaluation?" Descriptive analyses of participant ratings characterize the distribution of group responses from each round and estimate changes in group and individual responses between rounds. As is standard in ExpertLens panels [19,20], Round Three data are analyzed to identify consensus decisions on policy effectiveness and implementability by applying the inter-percentile range adjusted for symmetry (IPRAS) analysis technique outlined in the RAND/UCLA Appropriateness Method user manual [8]. IPRAS first determines whether disagreement exists among participants using calculated "disagreement index" scores, which indicate a lack of consensus on a policy for a given effectiveness outcome or implementation criterion (see Box). If no disagreement exists, Round Three medians determine consensus expert opinions. Sensitivity analyses explore robustness using Round One ratings to impute Round Three responses for participants who did not complete Round Three. As a secondary analysis, for the effectiveness panel, within-participant (between-rating) correlations assess whether expert ratings are concordant with the theory of change. Results from our correlational analyses validated that participants appeared to be providing internally consistent ratings and provided suggestive evidence on the extent to which participants viewed outputs or proximate outcomes as important mediators of more distal outcomes (e.g., the importance of treatment retention for opioid overdose mortality reduction).

Qualitative Data Analysis. Our qualitative data analysis seeks to answer the question: "what are expert views on the effectiveness and implementability of the types of policy under evaluation?" Reflexive thematic analysis on the entire qualitative dataset (i.e., all comments from the two rating rounds and discussion round) examines expert views on policy effectiveness and implementability. Reflexive thematic analysis is a flexible process of data familiarization, data coding, summarizing topics, and developing, revising, and finalizing themes [21]. While its phases are sequential and build off previous phases, it is typically a recursive process that
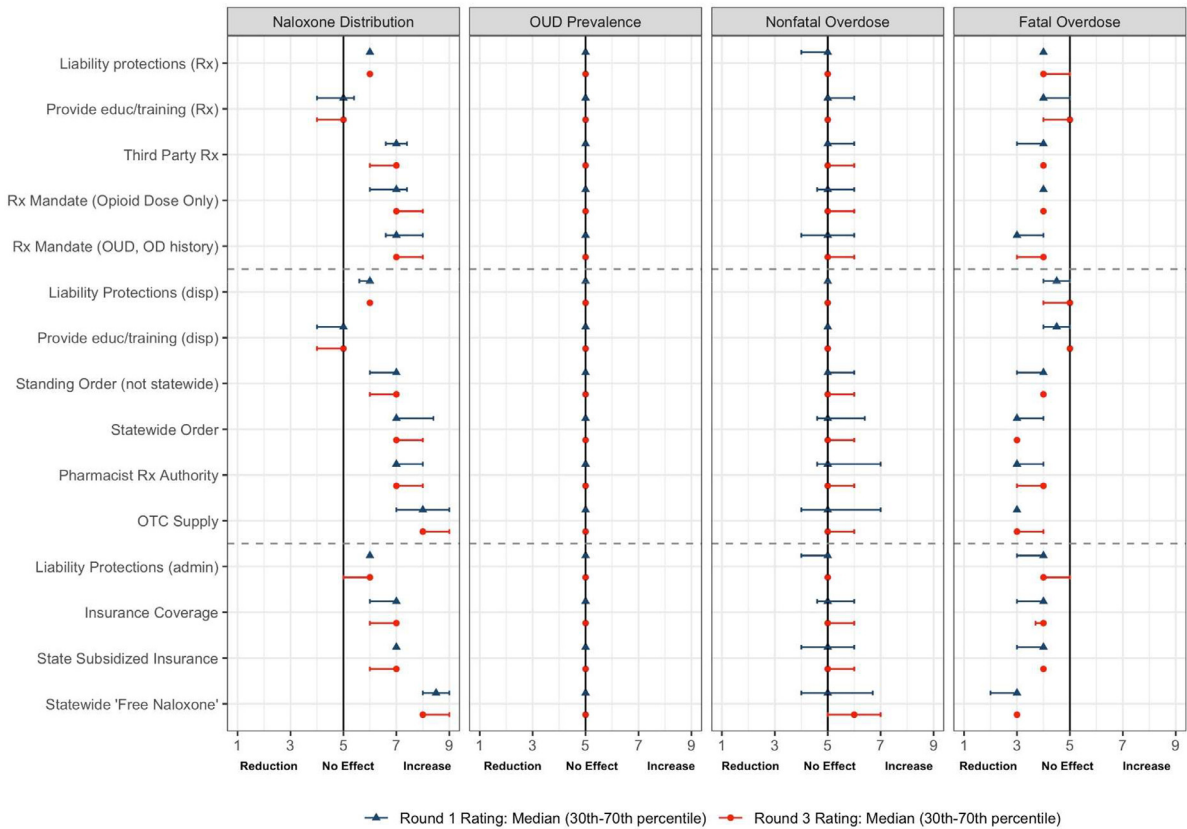
Fig. 3. Example visualization of effectiveness panel ratings.

Notes: OUD = opioid use disorder; OD = overdose; OTC = over-the-counter. Markers represent the median rating, with bars representing the inter-percentile range from the 30th to 70th percentile. Each panelist (n = 23 for Round 1; n = 21 for Round 3) rated items on a scale of 1 to 9, with scores of 1 to 4 indicating the policy leads to reductions in the outcome, a score of 5 indicating no effect, and scores of 6 to 9 indicating the policy leads to increases in the outcome.

involves movement back and forth between different phases. The first phase involves grouping all comments by policy and criterion, ordering comments within these groupings by experts' ratings, and thoroughly reading and re-reading the comments to become immersed and intimately familiar with their content. After this process of data organization and familiarization, the next phase is to conduct comment-by-comment coding, using succinct labels (codes) that are grounded in the data itself and might be relevant to addressing the research question. The third phase is to collate codes and examine data to summarize expert comments about each effectiveness and implementability criterion. The final phase is to create themes that capture significant, broader patterns of meaning across the dataset. This qualitative data analysis approach ultimately yields summaries of expert comments for each policy and rating criterion, as well as inferred themes about which policies are high-versus-low value: i.e., which to consider implementing-versus-deimplementing, sub-group considerations, and implementation considerations [1]. In our initial applications of ROMPER, one member of the research team conducted the analysis, while a second member of the team familiarized themselves with the data to check how the first researcher coded the data, clarify understanding of themes, and identify nuances and insights potentially missing from the analysis.

Presenting results

ROMPER involves presenting results visually and narratively. Rating results are visualized using a forest plot format (see Figs. 3 and 4). Similar to points for effect estimates and lines for confidence intervals, the median rating is represented as a point and the inter-percentile range is represented as a horizontal line extending either side of the point [22]. While guidelines for reporting Delphi processes are still forthcoming [23], ACCORD (ACcurate COnsensus Reporting Document) should be followed where appropriate [24], and the Standards for Reporting Qualitative Research should be followed to narratively report qualitative findings [25].

Given the end goal of ROMPER is to provide actionable information for decisionmakers who may be consider a menu of policy options, "policy profiles" are a means of conveying the online Delphi results in a manner tailored for a policymaker audience (see Fig. 5). Each policy profile focuses on a particular policy on which experts are queried, providing a GRADEPro Evidence-to-Decision
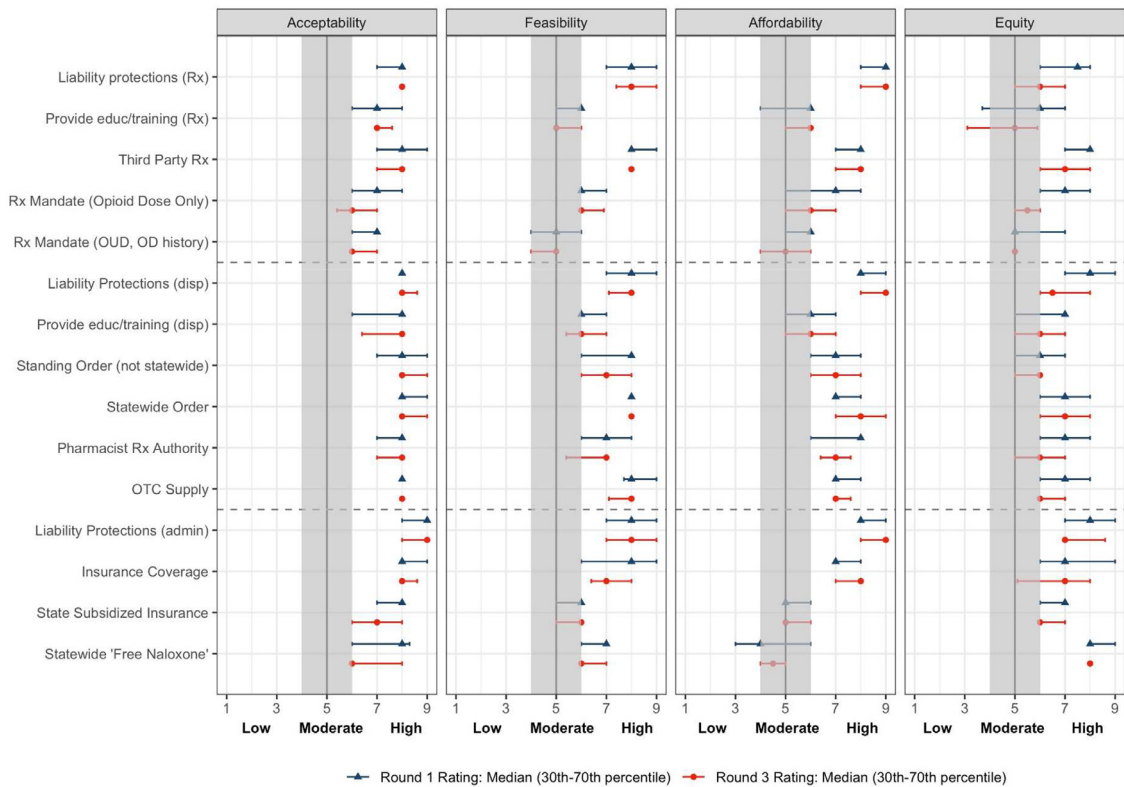
Fig. 4. Example visualization of implementability panel ratings.
Notes: OUD = opioid use disorder; OD = overdose; OTC = over-the-counter. Markers indicate the median rating. Bars indicate the inter-percentile range from the 30th to 70th percentile. Each panelist ($n = 21$ for Round 1; $n = 19$ for Round 3) rated items on a scale of 1 to 9, with scores of 1 to 3 indicating "low" acceptability, feasibility, affordability or equitability; scores of 4 to 6 indicating "moderate" acceptability, feasibility, affordability, or equitability; and scores of 7 to 9 indicating "high" acceptability, feasibility, affordability, or equitability.

table to summarize views on effectiveness and implementability as well as summary contextual information based on the qualitative analyses (https://osf.io/gb936/).

Method validation

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, and 7 = strongly agree) to rate 23 close-ended statements about their experiences with the Delphi study (i.e. ROMPER), online discussions, and the ExpertLens online system [26–28]. We combined data across panels and calculated univariate descriptive statistics for each item. We used original labels on the 7-point scale to describe expert feedback (i.e., median values and percentage of experts who gave positive, negative, or neutral ratings). Overall, experts generally reported positive experiences with ROMPER, the online discussion, and the ExpertLens platform (see Table 1). Regarding ROMPER, experts strongly agreed that the topic of the panels were important (Median [Med] = 7, Interquartile Range [IQR] = 7 to 7) and their expertise/experience were relevant to the topics of this panels (Med 7, IQR 6–7). In addition, most agreed that the panels would generate useful outcomes (Med 6, IQR 5–6.5), participating in the panels was satisfying (Med 6, IQR 5–6), and participation in the panels met their expectations (Med 6, IQR 5–6). However, most slightly agreed that participation in the panels took a lot of effort (Med 5, IQR 4–6). Regarding online discussions, most experts agreed that other participants in the discussion were respectful (Med 6, IQR 6–7), they were comfortable sharing their views (Med 6, IQR 6–7), the charts of Round 1 ratings were clear (Med 6, IQR 5–7) and helped them understand how their responses compared to those of other participants (Med 6, IQR 6–7), and the discussion was informative (Med 6, IQR 4–6). They only slightly agreed that the discussion digests were easy to understand (Med 5, IQR 4–6) and motivated them to post new discussion comments (Med 5, IQR 4–6), and that the discussion in Round 2 influenced their Round 3 ratings (Med 5, IQR 5–6). They also slightly disagreed that the digests were sent too frequently (Med 3, IQR 2–4) and that they had trouble following discussions (Med 3, IQR 2–5). However, they slightly agreed it was tedious to complete Round Two (Med 5, IQR 4–6). They had neutral views about panel moderators (Med 4, IQR 4–5) and feeling overloaded with information during the discussion (Med 4, IQR 2–5). Regarding ExpertLens, experts agreed that they were able to express their views on the study topic using the ExpertLens process (Med 6, IQR 6–7) and that the ExpertLens system was easy to use (Med 6, IQR 5–6). They only slightly agreed that they would like to use ExpertLens in the

Fig. 5. Example policy profile.

future (Med 5, IQR 4–6) and slightly disagreed that the mechanics of participating in ExpertLens distract from the substance of study (Med 3, IQR 2–5).

We also analyzed comments experts provided on how they found the overall experience and how to improve it in the future (see Table 2). First, experts emphasized the importance of considering fatigue—and countering it by keeping the scope manageable and avoiding unnecessary redundancy/repetitiveness in policies and rating criteria. Relatedly, experts made numerous comments about making the online system easy to navigate and user-friendly. Specific recommendations included sending frequent reminders during the discussion round and utilizing the moderator in the discussion forum to engage participants. A full list of policies upfront with crystal-clear definitions was also desired. Experts also expressed that it is useful to clearly and repeatedly communicate the amount of time and effort involved in participating in the full panel so that they can best plan accordingly. Many experts provided comments related to the utility of anonymous, controlled feedback about the responses of their fellow panelists. However, several experts also commented that they would be interested in adding a synchronous component for engaging participants in-person or via video conferencing. A few also wanted more opportunity to shape the design of the panel itself (e.g., policies and criteria). Lastly, experts provided suggestions for further information that would be useful: namely, summaries of existing research evidence on the policies, as well as more details about who else was participating in the panel (while still maintaining anonymity).

Limitations

We offer some self-reflections based on these data. First, to reduce participant burden, we limited the number of policies in each panel and experts to participating only in one panel. These choices consequently meant limiting the scope of our panels and the

Table 1

Expert responses to feedback questionnaire.

Item	N	Median	IQR-L	IQR-U	1–3	4	5–7
ROMPER							
Participation in this study was satisfying	88	6	5	6	6 %	17 %	77 %
The topic of this study is important	87	7	7	7	0 %	0 %	100 %
The study will generate useful outcomes	87	6	5	6.5	6 %	3 %	91 %
Participation in this study met my expectations	85	6	5	6	4 %	19 %	78 %
*Participation in this study took a lot of effort	86	5	4	6	22 %	16 %	62 %
My expertise/experience is relevant to the topic of this study	86	7	6	7	1 %	0 %	99 %
Online Discussion							
*It was tedious to complete Round Two	85	5	4	6	24 %	18 %	59 %
*I had trouble following discussions	86	3	2	5	55 %	15 %	30 %
The moderator led the discussion well	86	4	4	5	12 %	50 %	38 %
The discussion was informative	84	6	4	6	7 %	21 %	71 %
Other participants in the discussion were respectful	87	6	6	7	0 %	5 %	95 %
I was comfortable sharing my views	85	6	6	7	1 %	6 %	93 %
*I felt overloaded with information during the discussion	86	4	2	5	47 %	22 %	31 %
The charts in Round Two were clear	85	6	5	7	6 %	9 %	85 %
The charts helped me understand how my responses compared to those of other participants	87	6	6	7	5 %	7 %	89 %
The discussion digests were easy to understand	87	5	4	6	13 %	14 %	74 %
The digests motivated me to post new discussion comments	86	5	4	6	23 %	22 %	55 %
*The digests were sent too frequently	86	3	2	4	58 %	33 %	9 %
The discussion in Round Two influenced my Round Three ratings	87	5	5	6	10 %	6 %	84 %
ExpertLens System							
*The mechanics of participating in ExpertLens distract from the substance of study	87	3	2	5	54 %	16 %	30 %
I was able to express my views on the study topic using the ExpertLens process	85	6	6	7	2 %	2 %	95 %
The ExpertLens system is easy to use	86	6	5	6	6 %	8 %	86 %
I would like to use ExpertLens in the future	84	5	4	6	6 %	37 %	57 %

Notes. *indicates item where disagreement (lower score) signals a positive experience.

Table 2

Expert Responses to Open-Ended Feedback Questions.

Feedback	Quotations
Be mindful of scope, redundancy, and repetitiveness to prevent fatigue	"Yes, the process was tedious. Perhaps because it covered too many issues." (B30, Treatment)
Communicate the amount of time and effort involved	"I underestimated the amount of time it would take and therefore was unable to complete everything by given deadlines, which made me feel guilty." (B05, Naloxone)
Consider synchronous engagement among participants	"I prefer live discussions with opportunity for back and forth between individuals with differing opinions. I have a harder time getting/staying engaged with platforms like this. I realize those logistics are much more difficult to maneuver." (B19, Treatment)
Make the online system easy to navigate and user-friendly	"It was difficult to remember sometimes what policy you were on. if it was tightened up in terms of spacing, and the policy was available with every prompt, it might be easier as you got further down the page." (A08, Naloxone)
Provide anonymous, controlled feedback	"I did appreciate blinded participation, seeing my fellow participant answers, and seeing my past answer, some summary from moderator. It's clear that on some points, there is a lot of consensus. Others had more variation and hopefully that will be instructive." (A15, Naloxone)
Provide experts opportunity to shape panel design	"I wish you would have asked us to design our own optimal policy, even if by checkboxes or something with limited fills." (B06, Naloxone)
Provide experts with summaries of existing research evidence	"Would prefer discussing and actual review of the literature. I got the sense that most participants extrapolated a lot from weak literature that supported their biases. This is a challenge in the absence of a systematic review of each of the policies that informs the discussion." (A27, Treatment)
Provide list of and clear definitions for policies and rating criteria	"I would list all the scenarios at the beginning. This would help you understand the range so you can better see future scenarios without having to go back to previous pages. For example, once I found the scenario that I thought would have the largest impact, I felt like I had to go back to previous pages to modify my responses to reflect that. Also, would help to have clarification initially on definitions. I misinterpreted the nonfatal OD question-scenarios that improve naloxone access would actually increase nonfatal OD (and decrease fatal OD), so I didn't answer correctly initially." (A14, Naloxone)
Provide more information about fellow panelists	"Fascinating to hear other folks perspectives on these issues. I would have liked to know more about the person's background to better understand their perspective." (B20, Treatment)
Send frequent reminders during the discussion round	"Real time alerts if someone responds to your comment directly." (B01, Naloxone)
Utilize a moderator in the discussion forum to engage participants	"Having the moderator pull out the topic that was interesting, provocative, challenging, common was very important and useful." (A08, Naloxone)

insightfulness of our panels. Second, our panels were US-based by design, so ROMPER has yet to be applied with an international panel of experts. In addition, while we aimed to capture a diversity of perspectives, both samples of experts in our panels identified as largely non-Hispanic white and researchers with expertise in empirical evidence rather than direct experience with policies. An online format also requires participants to have stable Internet access, proficiency with online survey systems, and several hours of

availability over the three rounds. Third, we used proprietary software (ExpertLens; GRADEPro) to design and conduct the panels; future work is needed to replicate the protocol in free-to-use and open source software. Lastly, while the nature of online panels has the potential to allow more perspectives to be captured, online formats may not allow experts to provide as much depth and nuance in their comments as they would have in a smaller, in-person setting.

Conclusion

ROMPER provides a validated, replicable, open access approach for eliciting expert views on policy effectiveness and implementability—and to summarize expert consensus for policymakers. ROMPER is amenable with prominent approaches for developing guidelines (i.e., GRADE and RAM) and can help identify potential high-value polices to adopt or low-value policies to de-implement, as well as important implementation and sub-group considerations.

Box. Inter-Percentile Range Adjusted for Symmetry (IPRAS) analysis technique [8]

Disagreement is indicated if IPR > 2.35 + 1.5 * (5 − ((IPR_L + IPR_U) / 2)), where IPR represents the 40 % inter-percentile range of the 30th (lower IPR, or IPR_L) to 70th (upper IPR, or IPR_U) percentile as follows:

1. Determine the lower limit of the 40 % inter-percentile range (IPR_L; 30th-percentile score)
2. Determine the upper limit of the 40 % inter-percentile range (IPR_U; 70th-percentile score)
3. Determine the central point of the IPR (IPR_C): (IPR_L + IPR_U)/2
4. Determine the "Asymmetry Index" (AI) on the 9-point Likert scale: 5 - IPR_C
5. Calculate the IPRAS for the item: 2.35 + (1.5*AI)
6. Calculate the IPR: IPR_U - IPR_L
7. Calculate the "Disagreement Index" (DI): IPR/IPRAS
8. Determine whether disagreement exists: DI > 1

If an item has disagreement (DI>1), it is considered to be uncertain on the rated effectiveness outcome or implementation criterion. If an item has agreement (DI≤1), the tertile in which the median rating for importance falls will be analyzed: a median score between 1 and 3 indicates either a decrease in the value of the outcome of interest (Effectiveness Panel) or being low in the implementation criterion (Implementation Panel); a median score between 4 and 6 indicates either a little-to-no change in the value of the outcome of interest (Effectiveness Panel) or being moderate in the implementation criterion (Implementation Panel); and a median score between 7 and 9 indicates either an increase in the value of the outcome of interest (Effectiveness Panel) or being high in the implementation criterion (Implementation Panel).

The bar chart represents the distribution of participants' responses regarding equitability for the example of pay-for-performance policies. The blue line indicates a group median. The red dot indicates a participant's own response. In this chart, the median rating was 5 on a 9-point scale; the participant's own response was 8, which is above the median value and outside of the inter-percentile range (blue shaded region). The table presents a summary of participants' comments from Round one, listing main reasons for providing low ratings [1–3], uncertain ratings [4–6], and high ratings [7–9]. At the bottom of the figure are links to the discussion threads with comments posted in round 2 about the equitability of the selected policy discussion threads with comments posted in Round Two about the equitability of this policy, along with the list of individual round 1 comments from each participant, sorted by the numeric ratings.

Ethics statements

Study procedures were given exempt status by the corresponding author's Institutional Review Board (Study 2018–0506).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

SG's spouse is a salaried employee of, and owns stock in, Eli Lilly and Company. SG has accompanied his spouse on company-sponsored travel. SG is a member of the RAND ExpertLens Team. RS declares no conflicts of interest.

CRediT authorship contribution statement

Sean Grant: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Rosanna Smart:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Acknowledgments

Supplementary material and/or additional information

ROMPER protocols, materials, data, and code are openly available with Creative Commons licensing (https://osf.io/gb936/).

References

[1] J. Moberg, A.D. Oxman, S. Rosenbaum, H.J. Schünemann, G. Guyatt, S. Flottorp, et al., The GRADE Evidence to Decision (EtD) framework for health system and public health decisions, Health Research Policy and Systems 16 (1) (2018) 45.

[2] S.M. Nutley, I. Walter, H.T. Davies, Using evidence: How research Can Inform Public Services, Bristol University Press, Bristol, 2007.

[3] H.J. Schünemann, Y. Zhang, A.D. Oxman, Distinguishing opinion from evidence in guidelines, BMJ 366 (2019) l4606.

[4] J. Jones, D. Hunter, Consensus methods for medical and health services research, BMJ 311 (7001) (1995) 376–380.

[5] G. Rowe, G. Wright, Expert opinions in forecasting: the role of the Delphi technique, in: JS Armstrong (Ed.), Principles of Forecasting: A Handbook For Researchers and Practitioners [Internet], Springer US, Boston, MA, 2001, pp. 125–144, doi:10.1007/978-0-306-47630-3_7. [cited 2023 Jul 13](International Series in Operations Research & Management Science). Available from.

[6] I. Belton, A. MacDonald, G. Wright, I. Hamlin, Improving the practical application of the Delphi method in group-based judgment: a six-step prescription for a well-founded and defensible process, Technol. Forecast. Soc. Change 147 (2019) 72–82.

[7] M.K. Murphy, N.A. Black, D.L. Lamping, C.M. McKee, C.F. Sanderson, J. Askham, et al., Consensus development methods, and their use in clinical guideline development, Health Technol. Assess. 2 (3) (1998) 1–88 i–iv.

[8] K. Fitch, S.J. Bernstein, M.D. Aguilar, B. Burnand, J.R. LaCalle, P. Lazaro, et al., The RAND/UCLA Appropriateness Method User's Manual, RAND Corporation, 2001 [Internet]Jan [cited 2022 Dec 19]. Available from https://www.rand.org/pubs/monograph_reports/MR1269.html .

[9] D. Khodyakov, S. Grant, B. Denger, K. Kinnett, A. Martin, M. Booth, et al., Using an online, modified Delphi approach to engage patients and caregivers in determining the patient-centeredness of duchenne muscular dystrophy care considerations, Med. Decis. Mak. 39 (8) (2019) 1019–1031.

[10] R. Smart, S. Grant, Effectiveness and implementability of state-level naloxone access policies: expert consensus from an online modified-Delphi process, Int. J. Drug Pol. 98 (2021) 103383.

[11] S. Grant, R. Smart, Expert views on state-level naloxone access laws: a qualitative analysis of an online modified-Delphi process, Harm. Reduct. J. 19 (1) (2022) 64.

[12] R. Smart, S. Grant, A.J. Gordon, R.L. Pacula, B.D. Stein, Expert panel consensus on state-level policies to improve engagement and retention in treatment for opioid use disorder, JAMA Health Forum 3 (9) (2022) e223285.

[13] S. Grant, R. Smart, A.J. Gordon, R.L. Pacula, B.D. Stein, Expert views on state policies to improve engagement and retention in treatment for opioid use disorder: a qualitative analysis of an online modified Delphi process, J. Addict. Med. 18 (2) (2024) 129.

[14] S. Dalal, D. Khodyakov, R. Srinivasan, S. Straus, J. Adams, ExpertLens: a system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge, Technol. Forecast. Soc. Change 78 (8) (2011) 1426–1444.

[15] P. Alonso-Coello, H.J. Schünemann, J. Moberg, R. Brignardello-Petersen, E.A. Akl, M. Davoli, et al., GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction, BMJ 353 (2016) i2016.

[16] S. Grant, R. Smart, B.D. Stein, We need a taxonomy of state-level opioid policies, JAMA Health Forum 1 (2) (2020) e200050.

[17] K. Pottie, O. Magwood, P. Rahman, T. Concannon, P. Alonso-Coello, A.J. Garcia, et al., GRADE Concept Paper 1: validating the "F.A.C.E" instrument using stakeholder perceptions of feasibility, acceptability, cost, and equity in guideline implement, J. Clin. Epidemiol. 131 (2021) 133–140.

[18] D. Khodyakov, S. Hempel, L. Rubenstein, P. Shekelle, R. Foy, S. Salem-Schatz, et al., Conducting Online Expert panels: a feasibility and experimental replicability study, BMC Med. Res. Methodol. 11 (1) (2011) 174.

[19] J.S. Merlin, D. Khodyakov, R. Arnold, H.W. Bulls, E. Dao, J. Kapo, et al., Expert panel consensus on management of advanced cancer–related pain in individuals with opioid use disorder, JAMA Netw. Open 4 (12) (2021) e2139968.

[20] T.R. Radomski, A. Decker, D. Khodyakov, C.T. Thorpe, J.T. Hanlon, M.S. Roberts, et al., Development of a metric to detect and decrease low-value prescribing in older adults, JAMA Netw. Open 5 (2) (2022) e2148599.

[21] V. Braun, V. Clarke, Thematic Analysis: A Practical Guide, Sage Publishing, 2021.

[22] J.J. Deeks, J. Higgins, D.G Altman, Chapter 10: analysing data and undertaking meta-analyses, Cochrane Handbook for Systematic Reviews of Interventions version 63, Cochrane, 2022 [Internet][cited 2023 Jul 21]. Available from https://training.cochrane.org/handbook/current/chapter 10.

[23] J. Spranger, A. Homberg, M. Sonnberger, M. Niederberger, Reporting guidelines for Delphi techniques in health sciences: a methodological review. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 172 (2022) 1–11.

[24] W.T. Gattrell, P. Logullo, Zuuren EJ van, A. Price, E.L. Hughes, P. Blazey, et al., ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi, PLoS Med. 21 (1) (2024) e1004326.

[25] B.C. O'Brien, I.B. Harris, T.J. Beckman, D.A. Reed, D.A. Cook, Standards for Reporting Qualitative Research: a Synthesis of Recommendations, Acad. Med. 89 (9) (2014) 1245.

[26] D. Khodyakov, S. Grant, C.E.H. Barber, D.A. Marshall, J.M. Esdaile, D. Lacaille, Acceptability of an online modified Delphi panel approach for developing health services performance measures: results from 3 panels on arthritis research, J. Eval. Clin. Pract. 23 (2) (2017) 354–360.

[27] D. Khodyakov, S. Grant, D. Meeker, M. Booth, N. Pacheco-Santivanez, K.K. Kim, Comparative analysis of stakeholder experiences with an online approach to prioritizing patient-centered research topics, J. Am. Med. Inform. Assoc. 24 (3) (2017) 537–543.

[28] C.K. Armstrong, S. Grant, K. Kinnett, B. Denger, A. Martin, I. Coulter, et al., Participant experiences with a new online modified-Delphi approach for engaging patients and caregivers in developing clinical guidelines, Eur. J. Pers. Cent. Healthc. 7 (3) (2019) 476–489.