

Prioritizing research needs based on a systematic evidence review: a pilot process for engaging stakeholders

Rachel Gold PhD, MPH,* Evelyn P. Whitlock MD, MPH,† Carrie D. Patnode PhD, MPH,* Paul S. McGinnis MPA,‡ David I. Buckley MD, MPH,§ and Cynthia Morris PhD, MPH¶

*Research Associate III, The Kaiser Permanente Center for Health Research, Portland, OR, †Senior Investigator, Director, Evidence-based Medicine; Associate Director, Oregon Evidence-based Practice Center, ‡Community Health, Quality and Practice Development Director, §Investigator Department of Family Medicine, OHSU, Portland, OR and ¶Research Director and Investigator, Department of Medical Informatics and Clinical Epidemiology, OHSU, Portland, OR, USA

Abstract

Correspondence

Rachel Gold PhD, MPH
Research Associate III
Kaiser Permanente Center for Health
Research
3800 N. Interstate Avenue
Portland
OR 97227
USA
E-mail: Rachel.Gold@kpchr.org

Accepted for publication

15 June 2011

Keywords: research prioritization, stakeholder involvement, systematic evidence review

Background/context Systematic evidence reviews (SERs) identify knowledge gaps in the literature, a logical starting place for prioritizing future research. Varied methods have been used to elicit diverse stakeholders' input in such prioritization.

Objective To pilot a simple, easily replicable process for simultaneously soliciting consumer, clinician and researcher input in the identification of research priorities, based on the results of the 2009 SER on screening adults for depression in primary care.

Methods We recruited 20 clinicians, clinic staff, researchers and patient advocates to participate in a half-day event in October 2009. We presented SER research methods and the results of the 2009 SER. Participants took part in focus groups, organized by profession; broad themes from these groups were then prioritized in a formal exercise. The focus group content was also subsequently analysed for specific themes.

Results Focus group themes generally reacted to the evidence presented; few were articulated as research questions. Themes included the need for resources to respond to positive depression screens, the impact of depression screening on delivery systems, concerns that screening tools do not address comorbid or situational causes of depression and a perceived 'disconnect' between screening and treatment. The two highest-priority themes were the system effects of screening for depression and whether depression screening effectively leads to improved treatment.

Conclusion We successfully piloted a simple, half-day, easily replicable multi-stakeholder engagement process based on the results of a recent SER. We recommend a number of potential improvements in future endeavours to replicate this process.

Background

Systematic evidence reviews (SERs), including effectiveness and comparative effectiveness reviews (CERs), use rigorous methods to evaluate the extent to which existing research answers specific clinical questions. These questions usually involve whether a particular aspect of preventive care or treatment (e.g. a screening test or an intervention) directly or indirectly affects its targeted morbidity and whether any harms or risks are known to be associated with such care. The SER process also identifies knowledge gaps in the literature, such as insufficient or poor quality evidence to answer a given question overall, or in specific populations; the need for specific research methods to address missing evidence (e.g. clinical trials to study specific outcomes or compare the effectiveness of different treatments or epidemiologic studies of the prevalence or harms of a given treatment); or analyses of factors relevant to implementing evidence-based interventions. These gaps present a logical starting place for the identification and prioritization of future research questions in a given clinical area.

Traditionally, the prioritization of research topics has been driven by ‘experts,’ funding availability and researchers.¹ Recently, however, there has been increasing emphasis on engaging diverse stakeholders, including practitioners and patients, in identifying research priorities, with the goal of identifying the questions most relevant to improving clinical practice. The American Recovery and Reinvestment Act of 2009 allocated substantial funds to engage stakeholders in evidence gap identification and prioritization.² One example is the ‘Citizens’ Forum’ aimed to expand citizen and stakeholder engagement in the Agency for Healthcare Research and Quality’s (AHRQ) comparative effectiveness research initiative. AHRQ is also engaging the Evidence-based Practice Centers (EPC) to pilot methods for identifying and prioritizing evidence needs and to explore methodological issues around engaging stakeholders.³

Previous approaches to eliciting patient, clinician and/or expert input in setting research

priorities used methods including focus groups, online surveys and ‘forums.’^{4–15} A few studies solicited input from mixed groups of stakeholders;^{16–19} for example, Owens *et al.*¹⁹ used Delphi procedures to engage diverse stakeholders in identifying research priorities related to mental health care. Chalkidou *et al.* (2009)¹⁸ sought to generate a list of research priorities based on a current CER. Opportunities and challenges are evident in all prior approaches.

The pilot project presented here sought to test a simple, easily replicable process for simultaneously soliciting consumer, clinician and researcher input in the identification of research priorities, based on SER results. We hoped to identify processes that could be implemented subsequent to any future SER. The 2009 US Preventive Services Task Force report on screening for depression among adults in primary care,²⁰ conducted by the Oregon EPC, was selected as the systematic review for this pilot project because it was completed recently and because of the high prevalence of depression in primary care.

Details of this SER are published.²⁰ In summary, the SER assessed the known outcomes of screening adults for depression in primary care settings, building on a 2002 SER which showed that many screening instruments can reliably identify depressed adults in primary care, and that screening can lead to increased identification and treatment of depression.²¹ The 2009 review added evidence of minimal risks associated with depression screening and that common treatments can improve depression, although antidepressant use is associated with some risks of adverse effects.²⁰ There was minimal, mixed evidence to support screening unless it is followed by intensive case/care management. The 2009 SER identified a number of evidence gaps. These included the need for (i) clinical trials with unscreened control groups and evaluation of response, remission and health risks; and better information on (ii) the most effective components of primary care depression management programmes; (iii) the frequency and severity of missed depression cases; (iv) indicators of self-harm risk; (v) the impact of genetics, socio-

demographic factors and medication interactions on antidepressant efficacy and risks; and (vi) adverse effects of depression care management approaches.

To our knowledge, this is one of the first studies attempting to engage diverse stakeholders simultaneously and in person to identify research priorities based on the results of a formal evidence review.¹⁸ We present an overview of the methods of the pilot project, describe the findings and discuss lessons learned.

Methods

Participants

We recruited primary care clinicians, clinical support staff, clinical researchers and representatives of patient advocacy groups. Outreach was conducted through group emails, personal emails, direct calls and snowball sampling. Participants were recruited through Oregon's practice-based research networks (Safety Net West and the Oregon Rural Practice-based Research Network); the Oregon Clinical and Translational Research Institute, which includes the Kaiser Permanente Northwest Center for Health Research and Oregon Health & Sciences University; and local mental health advocacy groups. All invited clinicians were asked to bring a member of their support staff, preferably a person who administered depression screenings. We offered Continuing Medical Education credits and reimbursement for lost clinical time and travel costs. All participants completed informed consent forms approved by the Kaiser Permanente Institutional Review Board.

Presentations

A half-day event with all participants was held in October 2009. The Associate Director of the Oregon EPC (EPW) explained that our goal was to test an approach for directly engaging stakeholders in reviewing SER results as the basis for identifying additional research needed to affect clinical decisions, using the adult

depression screening SER as an example. She then presented a half-hour overview of evidence-based medicine and SER research methods. Next, one of the 2009 SER's co-authors gave a 40-min overview of the SER's results; the SER identified a number of evidence gaps, as outlined above, but we did *not* describe these in detail, to avoid biasing responses in the subsequent discussions.

Focus groups

Following the background presentations, participants were separated into three focus groups. We grouped participants by profession, hoping that being among peers would help participants speak more freely, and because we expected that each group would bring different perspectives. Clinicians were in one group and clinical support staff in another. Researchers and patient advocates were placed together in a third group, as neither are clinical staff, and because there were fewer participants in those categories. Each focus group lasted 60–70 min, was led by an experienced moderator and was structured around broad, open-ended questions. Table 1 lists the questions that the group leaders were instructed to ask; other topics were discussed as driven by the conversation. A member of the research team observed each group and took notes on the general concepts raised; these notes informed the structured group discussion and prioritization activity that followed. Participants were reminded that the discussions were being recorded, but that their comments would remain anonymous.

Group discussion and prioritization activity

While the participants regrouped for lunch, the note takers and moderators met for half an hour in a separate room and summarized the broad themes identified from each group. The summarized themes were then presented back to the entire group orally and written on poster board, as the basis for the large group discussion (the list presented back to the participants is given verbatim in Table 2). This discussion was

Table 1 Focus group questions

-
1. You just saw a presentation about the systematic evidence review on screening for depression among adults in primary care. What did you think of it?
 2. How is the evidence we presented today on depression screening in adults applicable to you and your work?
 3. What evidence would you like to have seen presented that would be important to inform your work?
 4. What do you think is missing from the evidence for depression screening in adults? What do you need to know to better serve your patients/population?
 5. What kinds of research questions would you like to see asked to address what's missing? How would the results of such research change your work?
 6. Are there any specific patient populations you would like to see involved in future questions on depression screening in adults?
 7. What are the most important outcomes that should be considered in future questions on depression screening in adults?
 8. What kinds of issues do you think might make it harder to answer the questions of interest to you?
-

intended to identify similarities and differences in the groups' responses and to identify concrete areas for future research. Next, we conducted a nominal group process – a prioritization exercise in which each participant, given three stickers colour-coded by focus group, was asked to place the stickers next to the ideas that they considered of highest priority. Participants were allowed to place as many of their stickers as they wished on a given theme.

Process evaluation

Last, participants were asked to complete an evaluation form. Likert-type questions were asked about the usefulness of the background presentations, the effectiveness and inclusiveness of the focus groups and group discussion, the effectiveness and relevance of the overall meeting and open-ended questions about what participants liked, did not like or would change about the event.

Analysis

Audiotapes of the focus groups and group discussion were transcribed verbatim. Transcripts were reviewed line-by-line and then thematically coded according to the themes that emerged, using Atlas.ti 5.0 qualitative software (Scientific Software Development, Berlin, Germany). A range of quotes is provided below, edited for readability and confidentiality. Responses to the Likert-type evaluation form questions were

analysed quantitatively, and those to open-ended questions were analysed qualitatively using content analysis techniques.

Results

Twenty individuals participated in the half-day event. There were nine clinician participants from diverse practice settings including urban safety net clinics, a private health maintenance organization, a private teaching hospital and rural community practices. The support staff focus group included six medical assistants. Five participants took part in the researcher/advocate (RA) focus group: two patient advocates from local mental health advocacy groups and three RAs with interest in mental health issues, one of whom was a physician. All attendees stayed until the end of the event.

Focus groups

Analyses of the focus group discussions, conducted after the event, identified the following themes. In general, the themes focused on reactions to the presented evidence for depression screening; few were articulated as specific research questions. The RAs' discussion, in particular, often diverged from the proscribed focus group questions. Nevertheless, useful themes emerged.

Both the clinicians and clinical support staff focused on the evidence showing the need for resources such as on-site support staff, medical

Table 2 Prioritization of research questions by focus group

Themes	Clinicians (n = 9)	Support staff (n = 6)	Researchers (n = 3)	Advocates (n = 2)	Total (n = 20)
Clinician focus group					
What are system effects (ER visits, workload, provider burden), as well as patient medical and functional outcomes?	11	6	0	1	18
What are more salient functional outcomes from screening and treatment?	2	1	0	0	3
Clarify/redefine 'medical home' roles, assign screening for depression by someone other than primary care physician – then physician troubleshoots/triages positive screens to appropriate care	0	3	0	0	3
Could depression screening in communities be more effective than in primary care?	1	0	1	0	2
Should screening be targeted and if so to whom and how?	0	0	1	1	2
If you are a primary care physician and get a positive depression screen, what to do next?	1	0	0	0	1
Depression screening may not be as important as other health issues, is harder to implement and should be targeted; but for whom?	0	1	0	0	1
Studies in comorbid and special populations; be sure tools address socioeconomic status and cultural groups	0	1	0	0	1
What are broader practice and system implications as well as practical approaches for implementing depression screening?	0	0	0	0	0
Can depression screening be isolated from other mental health and medical issues?	0	0	0	0	0
What are barriers to screening and strategies to overcome there?	0	0	0	0	0
What are non-medication depression treatments? Is cognitive-behavioural psychotherapy really feasible?	0	0	0	0	0
Are there important subgroups or populations in whom to screen because case-finding is failing (e.g. depression manifests differently or is under-reported)?	1	1	0	0	2
What results (types of dx) from depression screening?	0	0	0	1	2
Would screening for a broader range of MH issues be more effective?	0	0	0	1	1
What are operational and logistical impacts of depression screening?	0	0	0	0	0
How much harm is associated with stigma and is it greater in certain cultures?	0	0	0	0	0
Need recommendations of which tools and what system (how to do it?)					
Support staff focus group					
Time restrictions are a problem; how best to work with or change this? We have a good screening tool and good support but not enough time; not enough time for follow-up	0	3	0	0	3
Does treatment for depression help our patients with other problems?	0	1	1	0	2
Current research does not address our patient population (homeless, rural, illiterate); 'In my patients screening is irrelevant'; Too many basic needs not being met; 'We know they are depressed'	0	1	0	0	1

Table 2 (Continued)

Themes	Support				Total (n = 20)
	Clinicians (n = 9)	staff (n = 6)	Researchers (n = 3)	Advocates (n = 2)	
Researcher/advocate focus group					
Spontaneous remission appears in some treatment groups – what have these people done? (Qualitative approach required)	1	1	0	0	2
Majority of patients appear to have left treatment; why? What have they done? Holistic, not reductionist	0	1	0	0	1
What does the case manager do that creates positive outcomes?	0	0	0	0	0
Screening: are tools good enough? → diagnosis → CHASM → treatment	1	5	2	3	11
Model wrong: Medical vs. psychosocial, chronic vs. acute → look at BOTH	0	1	0	1	2
Harms: narrow and medically defined – suicide, ideation; stigmatization; BROADEN	0	0	0	1	1
Can we ask focused questions and expect to get answers of any value? Must be holistic – prior training, substance abuse	0	0	0	0	0
Outcomes: QOL etc.; thinking of depression as disease, not psychosocial entity	0	0	0	0	0

care teams or medical home models, behavioural health specialists and adequate time and resources available to respond appropriately, for depression screening to be effective. For instance:

Clinician (C): [The evidence] shows you need to have the support system in place. You need to have a behavioral health person in every office. But how are you going to do that?

Support Staff (S): Maybe [screening] could be helpful if you have had ... more time ... and you could speak with your patient. It all goes back to time – do we have the time that we need to give our patients quality care?

The support staff group voiced related concerns about patients' ability to pay for follow-up care, if the clinic cannot provide it:

S: I feel unfortunate in [rural practice] because we don't have that many resources. I would love to refer [my patients] to a counselor but how much is that going to cost? [They say] I can't afford it, and that is the end of that. It's a dead end when you don't have anyone you can refer them to.

The clinician group discussed the importance of understanding the impact of depression

screening on system-level outcomes including workflow, resources and provider satisfaction – for example, the potential impact of depression screening if it results in a patient needing prolonged attention.

C: What is the impact on the system? On the provider? You have to look at other workload issues, the time the PA spends in the room, the time the provider spends in the room. What's the real life day-to-day work impact of screening?

C: It feels horrible to see ... patients who have been waiting while you have been with a crying patient, to say "I'm sorry I'm an hour late, I'm sorry." So you have ... a vested interest ... in not screening, because you know "I'm going to unleash the flood gates."

Another theme emphasized by both clinicians and support staff was that the screening tools reviewed in the SER did not take into account the bigger issues in their patients' lives, to the extent that depression screening was almost irrelevant in their practices.

S: Nine times out of ten we know they are depressed because of their environment and their needs not being met. A majority of [my patients] can't even eat a meal for the day, or don't have a

warm place to sleep. It is ineffective to screen ... when they are not stable enough to maintain treatment.

C: I work in a homeless clinic with a high prevalence of addiction. I don't know if screening for depression is applicable if you got beat up yesterday or you're in withdrawal or you slept on the church steps.

In a similar vein, the RAs noted that depression is often a symptom of other factors and that the screening tools included in the SER do not address these causes, affecting their diagnostic reliability. Members of this group saw this as a conflict between a 'biomedical' model vs. a more holistic approach:

Researcher/Advocate (RA): In our world, sometimes [depression] is a condition but most of the time it is a symptom of something else.

RA: I may be sad ... if I am being hit at home, or ... have another condition that looks like depression.

RA: There [are] a lot of things that depression is comorbid with, and yet [the presented research] really doesn't get at any of those issues at all.

The RAs also voiced concern that the screening tools generalize too broadly that there are various forms and causes of depression and diagnostic inaccuracy would affect the provision of appropriate treatment.

RA: These scales tend to lump all depression into the same category, [but] you can have exogenous depression, endogenous depression, a lifetime of trouble, turmoil ... The odds that anti-depressants will help for that case is not going to be all that high.

When asked directly about populations on whom they would like to see more research, support staff voiced interest in younger (i.e. 20- to 35-year-olds) populations, refugees/immigrants, people with chronic pain, people of varying socioeconomic backgrounds and those with a family history of depression. The RAs also saw a need for research with ethnically diverse populations. In a related theme, support staff participants noted that patients may feel anxious or embarrassed about questions related to mental health and may not

report their feelings or behaviours because of cultural beliefs.

All three groups identified additional outcomes that would be important to study in relation to depression screening. In addition to the impact on the health care system, outcomes included the impact of depression screening on patients' physical pain, happiness, ability to work or form relationships, activity level, substance use and management of co-morbidities and on longer-term outcomes than those addressed in the SER (e.g. depression remission and recovery rates).

C: The piece that ... I didn't hear presented ... is [the impact of screening on] peoples' ability to function, maintain activities of daily life, hold down a job, take care of their family or themselves.

C: I think [research should include the impact of depression screening on] comorbidity, chronic pain, substance abuse, post-traumatic stress disorder, misdiagnosed bipolar.

The clinician group perceived a disconnection between research results and what happens in practice, including who is screened for depression, how and why they are screened, what happens with a 'positive screen' and the general usefulness of screening. Some participants felt that the SER did not clearly support the need for systematic depression screening in primary care. One theme was the apparently missing link between screening, diagnosis and treatment. Some clinicians wanted more information about what screening tools are best, their feasibility in practice and whether they improve on clinician expertise. The RAs wanted similar evidence that proper diagnoses or care follows from depression screening results and expressed concern that the SER was not structured to specifically answer whether care after screening is appropriate; they saw a 'chasm' or 'black box' between screening and treatment outcomes.

RA: Is there a procedure in place once that flag goes up? That to me is the biggest chasm. All this research rushed to the end and they didn't hook it up in the middle.

RA: I want to know what's inside the black box. You can screen and then on the end of the black box, you've got treatment ... I want to know which therapies you looked at.

C: I'm not convinced that it makes much difference if we screen or we don't screen ... There's a statistical difference, but on an outcome level I'm not sure it matters.

Group discussion and prioritization activity

Project staff presented the main themes identified in the focus groups, including participants' responses to the SER results, and specific ideas for needed research. Table 2 presents the 30 themes that were presented, verbatim to how they were written on the poster board and the level of emphasis given to each in the prioritization exercise. Twenty of the 30 themes received prioritization 'dots;' two stood out. The first, identified by the clinician group and highly selected by the clinicians and support staff, addressed the system effects of widespread screening for depression in primary care – what would the impact be on workload, provider and system burden, and patient outcomes? The next highest prioritized theme came from the RAs and asked whether depression screening effectively leads to improved treatment. Other prioritized themes included how to incorporate depression screening and follow-up on positive screens given practices' time restrictions; the role of 'medical homes' and the possibility that staff other than clinicians could conduct depression screening; and a question about functional outcomes from depression screening and pharmacological treatment.

Process evaluation

All 20 participants completed the evaluation form; 17 identified their participant group. Based on a Likert scale (1 = strongly disagree, 5 = strongly agree), the average rating of the usefulness of the background presentations was 3.85; this was slightly higher among clinicians. The focus groups received higher ratings, from 4.35 to 4.40; scores were slightly higher among

the support staff. The group discussion/prioritization exercise ratings ranged from 3.80 to 4.25, slightly higher among the support staff and slightly lower among the advocates. Scores for the usefulness of the overall meeting ranged from 4.00 to 4.50, with lower scores from the advocates. In the open-ended questions, many noted enjoying the opportunity to meet with others in their field and to discuss a topic of interest to them. Concerns about the process included the division of focus groups by profession, the relevance of the SER results to their populations and a fear that nothing tangible would result from their participation.

Discussion

To our knowledge, this is one of the first studies to systematically engage stakeholders in a process for identifying future research needs/priorities, based on the results of an SER.^{3,18} It is one of just a few that involved diverse stakeholders simultaneously^{16–19} in an in-person process. The piloted process was able to successfully integrate multiple viewpoints on research priorities based on a recent SER, in a half-day process that could be easily repeated. We would, however, recommend a number of improvements in future endeavours to replicate the process.

Lessons learned

Participant recruitment and preparation

We invited mental health researchers hoping that they would help direct the discussions towards the identification of specific research question. However, we found that – particularly in the focus group setting – the researchers either represented themselves in an advocate role or acted more as observers of the process. Future efforts should better clarify the researchers' role.

We invited mental health patient advocates, rather than patients, for easier identification and because we were concerned about confidentiality and potential harms of involving mental health patients. In future efforts, recruitment from a general patient pool might yield more represen-

tative participants. As researchers and advocates were difficult to recruit, and we thus had fewer such participants, we placed them together in one focus group. The transcripts showed that the advocates gave far more input than the researchers; separate focus groups might have yielded more from the researchers.

Clinicians who worked in indigent populations, or clinics with few services for follow-up on positive depression screenings, said the SER results were not relevant to their practices. While this suggested subpopulations where further research is needed, it limited our ability to identify research priorities based on responses to this SER. An SER is, by design, limited to research conducted in pre-determined settings, which seeks to answer specific questions; it is also limited by what has been studied. Therefore, the piloted process might be improved through consideration of how well the participants fit the parameters set by the analytic framework and the available research. However, our finding that the SER was structured in a way that was not credible to certain participants, or relevant to certain practice settings, might have been lost with participants from more homogeneous practices. To address credibility, earlier involvement of stakeholders may be needed during SER development. Such an approach is now being implemented in AHRQ's efforts to involve stakeholders in SER development and refinement through their Effective Healthcare Program. While integrating diverse input into SER development might complicate the SER process, it could lead to SERs that ask more relevant questions. In a related vein, given participants' concerns that engaging in this exercise would have no impact, future efforts should include a clear explanation of how participants' input will be used. If feasible, a process for updating participants should be determined *a priori* and shared to alleviate these concerns and explain how their input could potentially be used.

Identifying research priorities through the piloted process

While research priorities could be inferred from the data collected, aspects of the piloted process made it challenging to identify specific research

questions suggested by the SER results. We successfully solicited diverse perspectives, but future efforts should take into account that diversity of participants means variability in comprehension of the background materials and the goals of the process. This concurs with Oliver and colleagues'¹⁴ conclusion that appropriate education of health care 'consumers' is needed to enable their participation in research topic prioritization. A 'one size fits all' presentation and questions might be reconsidered when recruiting participants with widely variable expertise. For example, the focus group questions asked about the relevance of the presented SER results to the participants' 'work.' This language may have been inappropriate for the patient advocates, affecting their ability to formulate research questions. Conversely, some participants critiqued our decision to divide the focus groups by profession; diversifying the focus groups would address this, but would complicate avoiding the 'one size fits all' approach. Our research team concluded that in future efforts, rather than directly asking participants to develop research questions, a better approach might be to intentionally engage indirectly – to present the evidence, let the participants discuss it freely, then extrapolate research ideas and priorities from the transcripts.

Identification of research priorities might also be facilitated by having the participants first establish a set of criteria to be used in the prioritization process. Menon and Stafinski⁵ engaged a 'citizen's jury' to identify and rank priorities for research on health technologies; the Effective Healthcare Program has priority criteria that they apply when selecting topics for evidence reviews.²²

Feedback from the participants – and our observations – highlighted the need to refine our background presentation. We summarized SER methodology in general and then described the specific parameters and findings of the SER in question, intending to present the SER results in context without biasing participants' discussion of research priorities. We found that our emphasis on the SER's parameters illuminated a

substantial disconnect between SER methods and participant perceptions of how primary care is or should be provided. Many participants were concerned that the SER's analytic framework – the pre-determined structure precisely defining an SER's questions, and which studies it includes – was based on a 'medical model', which presumes that screening leads to effective treatment. This generated the second-most prioritized theme: the need to bridge a perceived 'chasm' or 'black box' between screening and treatment. The analytic framework's focus on primary care settings was criticized as being relevant only to primary care providers and patients with access to primary care; this concern was reiterated in the process evaluation. While the SER's scope could not be addressed in this process, this highlights the need to engage diverse stakeholders in the scoping phase of the review process to ensure that relevant settings, populations and approaches are considered.

We concluded that to improve this process, both the presentation of the SER results, and the methods for soliciting the identification of research priorities, should be structured to be more accessible to a diverse group of stakeholders. Focus group questions should be more open-ended (e.g. 'What is your experience with screening for depression?') and/or more closely tailored to the SER-identified gaps (e.g. 'What would you need to know to determine whether or not to routinely screen your patients for depression?'), to address the difficulty we encountered in guiding participants towards identifying research gaps. Future efforts should also consider establishing specific criteria that participants can use to help prioritize research questions/gaps.

Group discussion/prioritization exercise

We felt that the project team had inadequate time during the half-day meeting in which to review their focus group notes and highlight main concepts. In future events, the research team should get enough time for this important step. Another option would be to allow the focus groups themselves to generate their list of

themes. Conversely, building the structured group discussion around team-identified themes made it more feasible to steer the discussion towards specific research needs; the prioritization exercise thus yielded the most research-specific results of the event. This exercise might have been improved had the focus group transcription text analyses been conducted prior to the prioritization exercise, so that the identified themes were based on the participants' words rather than team members' notes; however, this would require a follow-up meeting, introducing a level of logistical complexity counter to our goal of piloting an easily replicable, in-person process.

Topic selection

We believe the process might have been smoother had we not piloted it with an SER on screening for depression, a topic that is both emotionally laden for patients and clinically challenging for practitioners.

Comparison to previous attempts to solicit stakeholder input in research topic prioritization

Overall, our findings about engaging stakeholders in research prioritization via the piloted process were similar to previous attempts in several ways. Others reported that soliciting research prioritization input – be it from community members or clinician/experts – can be challenging if these key informants lack a basic understanding of the SER and clinical research processes.^{10,15} Despite the participants' differences in knowledge about research methods, we were able to solicit feedback that could be used to infer research priorities and found several important parallels between the participant groups in these priorities. This divergence in understanding of the research process but similarity in identified priorities is comparable with that reported by Owens *et al.*¹⁹

Some of the identified priorities paralleled the research 'gaps' identified in the SER, and some diverged. The SER-identified need for research on the impact of depression screening

on improved care and clinically relevant outcomes and on the primary care resources needed to respond to positive screens, paralleled our participants' interests. The more clinical, epidemiologic and methods-specific SER-identified 'gaps' were not reflected in our participants' priorities. The participant-identified need for research in more diverse primary care populations, and questions about the utility of screening for depression when environmental causes are not addressed, diverged from those identified by the SER. Dissimilarities between 'expert'-identified gaps and stakeholder-identified gaps have previously been noted.^{1,12}

In the most closely related previous attempt to solicit diverse stakeholder input in research prioritization based on SER results, Chalkidou *et al.*¹⁸ recruited a group of stakeholders, including clinicians, patients, researchers, hospital administrators, payers and representatives of relevant government and industry agencies to 'score' a set of research questions related to management of coronary artery disease. Their process included three meetings where stakeholders were asked to identify specific research questions (with the CER as a starting point, but allowing input from other sources), develop a list of prioritization criteria and then 'score' these questions according to the criteria. While this process successfully generated a list of prioritized research questions, it repeated some of the drawbacks of earlier attempts to engage stakeholders in research prioritization. Like most previous efforts to engage stakeholders in identifying research priorities, their multi-step procedures required participants to return on several occasions; this might be too costly and complex to be feasibly implemented following all SERs/CERs. Further, their process began with discussions of the CER's identified evidence gaps and involved much expert input – as well as industry input – in the identification of the research needs considered in their prioritization process, potentially biasing the prioritization exercise results away from the research questions most essential to informing clinical practice and patient-centred outcomes.

Limitations

We intended to pilot an easily replicable process in which community, clinician and researcher input was solicited to identify and prioritize future research needs, based on evidence from a recent SER. Several important limitations apply, some of which are inherent to focus group methods. First, while all of the focus group moderators were experienced in that role, their levels of experience varied, as did their expertise in the content area. All of the focus group leaders were directed to guide their groups through the questions in Table 1, but they had mixed success in doing so, and the content and level of detail of the groups' discussions varied accordingly. Second, in any focus group, there is a risk that strong personalities could sway the discussion and influence it unduly; the transcriptions from the focus groups, however, did not suggest that this occurred. Third, participants were not selected randomly. In particular, patient advocates are not representative of the general patient population, and the participant clinicians represented just certain kinds of practices; for example, no local private practitioners participated.

Conclusions

SERs are essentially the gold standard in assessing the evidence supporting facets of clinical practice, and identifying what evidence is missing. The opportunity to use their results to stimulate high-priority research is compelling. Concurrently, there is increasing emphasis on engaging stakeholders in research prioritization, with the goal of identifying the research questions that are the most important for clinical decision-making, and most important to the community. Hence, there is a pressing need to develop inexpensive, easily reproducible methods for the engagement of stakeholders in research prioritization based on SER-identified research gaps. Despite limitations, this project achieved its primary goals of piloting a simple, multi-stakeholder engagement process based on the results of a recent SER and providing useful information about how to improve methods for

soliciting stakeholder input in the future.

Acknowledgements

The authors thank Debra Burch, Michelle Eder, Arwen Bunce and Cheryl Johnson, for their assistance with this project. The authors also thank the clinical staff, patient advocates and researchers who participated in this research.

Funding and conflicts of interest

This research was funded by a grant from the Oregon Clinical and Translational Research Institute, grant number UL1 RR024140 01 from the National Center for Research Resources, a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. We have no financial conflicts of interest to disclose.

References

- 1 Tong A, Sainsbury P, Carter SM *et al.* Patients' priorities for health research: focus group study of patients with chronic kidney disease. *Nephrology, Dialysis, Transplantation*, 2008; **23**: 3206–3214.
- 2 Department of Health and Human Services. Federal Coordinating Council for Comparative effectiveness research. <http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf>, accessed 5 July 2011.
- 3 Agency for Healthcare Research and Quality. Future research needs – methods research series. Available at: <http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?page-action=displayProduct&productID=481>, accessed 1 April 2011.
- 4 Smith N, Mitton C, Peacock S, Cornelissen E, MacLeod S. Identifying research priorities for health care priority setting: a collaborative effort between managers and researchers. *BMC Health Services Research*, 2009; **9**: 165.
- 5 Menon D, Stafinski T. Engaging the public in priority-setting for health technology assessment: findings from a citizens' jury. *Health Expectations*, 2008; **11**: 282–293.
- 6 Goberman-Hill R, Horwood J, Calnan M. Citizens' juries in planning research priorities: process, engagement and outcome. *Health Expectations*, 2008; **11**: 272–281.
- 7 Entwistle V, Calnan M, Dieppe P. Consumer involvement in setting the health services research agenda: persistent questions of value. *Journal of Health Services Research & Policy*, 2008; **13** (Suppl. 3): 76–81.
- 8 Perkins P, Booth S, Vowler SL, Barclay S. What are patients' priorities for palliative care research? – A questionnaire study. *Palliative Medicine*, 2008; **22**: 7–12.
- 9 Henschke N, Maher CG, Refshauge KM, Das A, McAuley JH. Low back pain research priorities: a survey of primary care practitioners. *BMC Family Practice*, 2007; **8**: 40.
- 10 Abma TA. Patients as partners in a health research agenda setting: the feasibility of a participatory methodology. *Evaluation and the Health Professions*, 2006; **29**: 424–439.
- 11 Welfare MR, Colligan J, Molyneux S, Pearson P, Barton JR. The identification of topics for research that are important to people with ulcerative colitis. *European Journal of Gastroenterology and Hepatology*, 2006; **18**: 939–944.
- 12 Brown K, Dyas J, Chahal P, Khalil Y, Riaz P, Cummings-Jones J. Discovering the research priorities of people with diabetes in a multicultural community: a focus group study. *British Journal of General Practice*, 2006; **56**: 206–213.
- 13 Oliver S. Patient involvement in setting research agendas. *European Journal of Gastroenterology and Hepatology*, 2006; **18**: 935–938.
- 14 Oliver S, Clarke-Jones L, Rees R *et al.* Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technology Assessment*, 2004; **8**: 1–148, III–IV.
- 15 Caron-Flinterman JF, Broerse JE, Teerling J, Bunders JF. Patients' priorities concerning health research: the case of asthma and COPD research in the Netherlands. *Health Expectations*, 2005; **8**: 253–263.
- 16 Malcolm C, Knighting K, Forbat L, Kearney N. Prioritization of future research topics for children's hospice care by its key stakeholders: a Delphi study. *Palliative Medicine*, 2009; **23**: 398–405.
- 17 McIntyre S, Novak I, Cusick A. Consensus research priorities for cerebral palsy: a Delphi survey of consumers, researchers, and clinicians. *Developmental Medicine and Child Neurology*, 2010; **52**: 270–275.
- 18 Chalkidou K, Whicher D, Kary W, Tunis S. Comparative effectiveness research priorities: identifying critical gaps in evidence for clinical and health policy decision making. *International Journal of Technology Assessment in Health Care*, 2009; **25**: 241–248.
- 19 Owens C, Ley A, Aitken P. Do different stakeholder groups share mental health research priorities? A four-arm Delphi study. *Health Expectations*, 2008; **11**: 418–431.
- 20 O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Annals of Internal Medicine*, 2009; **151**: 793–803.

- 21 Pignone MP, Gaynes BN, Rushton JL *et al.* Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 2002; **136**: 765–776.
- 22 Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *Journal of Clinical Epidemiology*, 2010; **63**: 491–501.