Living Systematic Reviews: Practical Considerations for the Agency for Healthcare Research and Quality Evidence-based Practice Center Program
White Paper
Living Systematic Reviews: Practical Considerations for the Agency for Healthcare Research and Quality Evidence-based Practice Center Program

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 75Q80120D00006

Prepared by:
Pacific Northwest Evidence-based Practice Center
Portland, OR

Investigators:
Marian S. McDonagh, Pharm.D.
Roger Chou, M.D.
Jesse Wagner, M.A.
Azrah Y. Ahmed, B.A.
Benjamin J. Morasco
Suchitra Iyer, Ph.D.
Devan Kansagara, M.D., M.C.R.
This report is based on research conducted by the Pacific Northwest Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 75Q80120D00006). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this White Paper, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Robert “Bob” Otto Valdez, Ph.D., M.H.S.A.
Director
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Craig A. Umscheid, M.D., M.S.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Suchitra Iyer, Ph.D.
Task Order Officer
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality
Acknowledgments

The authors gratefully acknowledge Task Order Officer Meghan Wagner, Pharm.D., M.B.A. at the Agency for Healthcare Research and Quality for contributions to this project.
Abstract

Living systematic reviews are a relatively new approach to keeping the evidence in systematic reviews current by frequent surveillance and updating. The Agency for Healthcare Research and Quality’s Evidence-based Practice Center Program recently commissioned a systematic review of plant-based treatments for chronic pain management. This white paper describes the team’s experience in implementing the protocol that was developed a priori, and reflects on the challenges faced and lessons learned in the process of developing and maintaining a living systematic review. Challenges related to scoping, conducting searches, selecting studies, abstracting data, assessing risk of bias, conducting meta-analysis, performing narrative synthesis, assessing strength of evidence, and generating conclusions are described, as well as potential approaches to addressing these challenges.
Contents

1. Background ................................................................................................................................. 1
   1.1 Case Study .......................................................................................................................... 1
   1.2 Considerations for a Living Systematic Review ............................................................ 1
2. AHRQ’s Experience ................................................................................................................... 2
   2.1 Interim Reports Before the Original Systematic Review Is Complete ....................... 2
   2.2 Scope ............................................................................................................................... 2
   2.3 Input From Experts .......................................................................................................... 3
   2.4 Literature Search ............................................................................................................. 3
   2.5 Study Screening, Selection, Abstraction ....................................................................... 4
   2.6 Updating Meta-analyses ................................................................................................. 4
   2.7 Updating Strength of Evidence Ratings ......................................................................... 5
   2.8 Peer Review of Interim Reports ..................................................................................... 5
   2.9 Surveillance Report Frequency ...................................................................................... 5
3. Conclusions ............................................................................................................................... 6
4. References ................................................................................................................................... 8

Tables
Table 1. Summary of literature searches and screening between reports ................................. 4
Table 2. Summary of LSR methods approaches, challenges, and proposed resolutions ........ 6
1. Background

Living systematic reviews (LSRs) are a relatively new approach of keeping the evidence in systematic reviews current by frequent surveillance and updating. The promise of up-to-date information is attractive to guideline developers, health system executives, and other end-users such as clinicians. Unlike for systematic reviews, standards for ‘living’ reviews are yet to be formalized, although guidance is available on many of the practical considerations.

As a prolific producer of systematic reviews in the United States with many private and public partnerships, the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center Program (EPC) has also been interested in better understanding methods, processes, and infrastructure needed to maintain living systematic reviews. Therefore, a topic with a potentially evolving evidence base presented an opportunity to pilot this approach and share our experience.

1.1 Case Study

AHRQ’s EPC Program recently commissioned a systematic review of plant-based treatments for chronic pain management. The purpose of the systematic review was to assess and synthesize the evidence on the benefits and harms of cannabis and other plant-based compounds (PBCs) as treatments for chronic pain.

Chronic pain is difficult to treat successfully, and a series of recent systematic reviews have found that commonly used treatments, including opioids, nonopioid medications, and nonpharmacological interventions, have limited efficacy. Opioids have received particular scrutiny due to their addiction and overdose potential, especially given the ongoing opioid crisis. Finding new, effective treatments for chronic pain is paramount, but it is unclear whether PBCs, such as cannabis and kratom, are effective and/or safe.

Given the trend towards state-level legalization of cannabis, and the lack of regulation of kratom and other PBCs, it is anticipated that research on their use for treating chronic pain will increase in coming months and years. AHRQ therefore funded a living systematic review spanning multiple years, to evaluate and update the evidence-base on these interventions as they get published.

1.2 Considerations for a Living Systematic Review

When beginning this review, we were aware of a number of potential challenges, detailed in a recent survey of participants in a LSR pilot project. This publication described (1) the need to develop methods, procedures and processes (automated if possible) for ongoing review; (2) the need to develop rules for when there is sufficient evidence to signal a change in conclusions; (3) the need to standardize how new evidence is reflected in updated review publications (version control); and (4) managing workload consistency and accuracy across time, including managing “burnout.”

This report summarizes the methods developed a priori to address the anticipated challenges and to improve efficiency in searching the literature, selecting studies, and making other methodologic decisions (https://effectivehealthcare.ahrq.gov/products/plant-based-chronic-pain-treatment/protocol; PROSPERO registration no. CRD42021229579). In addition, this report outlines several adjustments applied to the methods initially proposed. Challenges along with lessons learned are also described in this document.
2. AHRQ’s Experience

2.1 Interim Reports Before the Original Systematic Review Is Complete

The starting point of a living systematic review is usually a recently completed review. However, in order to be responsive to stakeholders, we created three progress reports (approximately every 3 months) while the original review on Cannabis for Chronic Pain Management was in progress, each describing studies identified during the interim period. The findings from all progress reports were synthesized and integrated into the ensuing systematic review, approximately 12 months after the start date. While this idea was successful in keeping stakeholders apprised of the emerging evidence, it posed a number of logistical challenges.

The development of the first interim report was the most time and labor intensive. A large bulk of the time and effort was taken up by this early stage since the output of the first search represented the greatest volume of literature to be screened. During this time, the format for the interim progress reports, data tables and summary of evidence tables also had to be developed concurrently. The first interim report was designed based on the early evidence collected. As additional studies included from the successive screenings prompted the review team to rethink their analytic approach, substantial modifications to the format of the second progress report were necessary (for example, creation of an organizational framework to categorize the cannabis interventions).

This front-loading of effort is an issue that needs consideration in the budgeting and staffing plans. While this is no longer an issue going forward for this review, future LSRs will likely have to plan for this if they plan to start with interim progress reports.

2.2 Scope

Several complexities of the available literature posed a challenge to operationalize the predefined inclusion-exclusion criteria (such as whether or not to include synthetically derived cannabinoids, which represent a large proportion of studies). Additional time in the early stage of the report development would have been particularly helpful as the review team engaged with the technical expert panel and considered how best to analyze and present the available information.

A “best evidence” approach was initially proposed, where observational studies would help fill gaps in evidence if randomized controlled trials were unavailable. For example, we included a study that was initially designed as a longitudinal study (single group) that changed methods part way through to include a comparison group and published interim results. Since there was no other evidence on the outcome of interest (effect on opioid use) for the type of intervention (medical cannabis program/patient’s choice of product), a decision to include this study was made. As better evidence (ideally from randomized controlled trials) emerges, we may eliminate this study from future analyses.

Many of the decisions about study inclusion made during the course of the initial systematic review may need to be adjusted as evidence changes and/or grows, or as new products/dosing strategies are introduced in the coming years. Therefore, the scope and protocol would benefit from periodic re-examination for potential revisions to keep the review relevant and based on the highest quality evidence. Future changes in scope may require additional decisions on how to best incorporate them. For example, should newly defined criteria be applied retrospectively,
would the review team be required to reconsider previously excluded studies, or would it suffice to prospectively apply new criteria only to future updates?

2.3 Input From Experts

For a standard AHRQ systematic review, a technical expert panel (TEP) made up of individuals with expertise in the topic is engaged early in the review process primarily to assist in refining the scope of the review. For an LSR it is useful to have a TEP that is engaged on an ongoing basis for the entire duration that the report is being updated, since additional scoping issues may arise over time. Our TEP had expertise in pharmacology, pain and addiction medicine, behavioral health, and prevention of problematic use. During our review, we obtained TEP input on several issues, such as whether to include new plant-based substances and whether to limit to randomized controlled trials, based on the number and quality of studies included. As we enter the phase of updating the systematic review, we are considering an expansion of the expertise represented on the TEP to match the complexities of the emerging literature. Particularly, a medicinal chemist may offer insights on the various cannabis products reported in the studies that may be extracted, purified, or isolated from the whole plant.

In addition to a TEP consisting of content experts, we envisioned engaging a group of experts who had experience with conducting and/or using LSR methods to provide input on our quarterly reports, processes, and methods. During the first year of this work (up to the completion of the original systematic review), we attempted to engage experts who had experience in LSRs and requested input on a progress report created to capture the interim state of the evidence between the draft and final systematic reviews. We sought feedback on the format and content of updates, methods, and language used to describe changes or stability of the strength of evidence (SOE), and their experience with automating parts of the review process. We received only limited input, primarily agreeing with our approach. The TEP noted some confusion with regard to the different report types (e.g., draft report vs. progress report). Obtaining input from external experts, beyond those on our own team, was difficult because LSRs are a new and evolving field with few people having adequate expertise to comment. We plan to continue to engage with the LSR experts as we produce future updates to help refine our methods and communication of the findings.

2.4 Literature Search

Regular, periodic updates of the literature searches are an essential underpinning of the LSR process that requires finding a balance between the frequency of search updates and efficiency in conducting and processing the search result. We used the same search strategy as the initial literature search throughout the first year. Starting in October 2020, automatic electronic searches using an “Auto-Alert” function were run every 2 weeks in Ovid (Medline, Cochrane Central, and PsycINFO), Embase and Scopus. These were staggered so that every week results of two or three databases were received for screening and moved through the review process. This frequency and the resulting smaller volumes of evidence were important in keeping the LSR on track. We were able to rapidly identify important new studies, especially those that might add to meta-analyses or change the SOE ratings.

There was some time lost when we discovered that 2 months’ worth of automatic search results were identified as spam by the institution’s email server and not delivered. The large volume of citations then had to be retrieved and screened manually. We had to make a special
effort to coordinate between the technology department, the library, and the database vendor to resolve this issue.

2.5 Study Screening, Selection, and Abstraction

Recognizing that frequent, iterative literature screening and study selection for LSRs can lead to reviewer-fatigue, we proposed to use systematic review software to screen abstracts. While the initial literature searches resulted in approximately 2,494 citations, potentially adequate to train the DistillerSR®’s AI feature, the review team ultimately decided to screen these initial abstracts manually. We conducted simultaneous dual review of abstracts (as usual, using DistillerSR®’s settings) to process these citations expeditiously. Once the review of abstracts and relevant full-text articles from the initial searches was complete, processing results from subsequent searches was less tedious, especially because the citation yield was significantly smaller (mean 7 citations per week, range 7 to 22). The team continued to employ a nearly simultaneous and manual dual review process to identify potentially relevant articles rapidly. The automation of abstract screening was ultimately not implemented.

Allotting adequate time to team members made it possible to efficiently make study inclusion decisions on a monthly basis. While the initial abstraction of the 17 studies for the first progress report was the most labor intensive, the volume of studies identified later were small and easily incorporated into the work flow.

Table 1 provides a summary of the time frame for literature searches and articles retrieved, screened, and included in each interim report and the systematic reviews.

### Table 1. Summary of literature searches and screening between reports

<table>
<thead>
<tr>
<th>Report</th>
<th>Literature Search End Date</th>
<th># New Abstracts Screened (Total Cumulative)</th>
<th># New Full-Text Reviews (Total Cumulative)</th>
<th>Included Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Report 1</td>
<td>Sept 14, 2020</td>
<td>2,494</td>
<td>148</td>
<td>17 (14 RCTS, 3 Observational)</td>
</tr>
<tr>
<td>Progress Report 2</td>
<td>Feb 1, 2021</td>
<td>138 (2,632)</td>
<td>39 (180)</td>
<td>6 RCTs</td>
</tr>
<tr>
<td>Draft SR</td>
<td>Mar 1, 2021</td>
<td>32 (2,671)</td>
<td>3 (183)</td>
<td>0 Studies</td>
</tr>
<tr>
<td>Progress Report 3</td>
<td>Apr 26, 2021</td>
<td>71 (2,742)</td>
<td>12 (195)</td>
<td>1 Observational</td>
</tr>
<tr>
<td>Final SR</td>
<td>Jul 16, 2021</td>
<td>108 (2,850)</td>
<td>19 (214)</td>
<td>3 Observational</td>
</tr>
<tr>
<td>Surveillance Report 1</td>
<td>Mid-Aug, 2021</td>
<td>32 (2,882)</td>
<td>3 (217)</td>
<td>0 Studies</td>
</tr>
<tr>
<td>Surveillance Report 2</td>
<td>Oct 29, 2021</td>
<td>102 (2,984)</td>
<td>27 (244)</td>
<td>1 RCT</td>
</tr>
</tbody>
</table>

Abbreviations: RCT = randomized controlled trial; SR = systematic review

2.6 Updating Meta-analyses

Continuous updating meta-analyses as new studies are incorporated into the report presents advantages and disadvantages. While an updated meta-analysis has the potential to improve effect size estimates as more data became available, there are several other issues to consider. First, the amount of dedicated effort from the statistician for the LSR process may need adjustment to ensure that adequate statistical support is available at multiple timepoints and on an ongoing basis, rather than late in the review process (as in a standard systematic review). However, this would mean setting aside statistician time without knowing if there will be evidence to meta-analyze.

As the evidence base matures, we will need to consider how often to update the meta-analyses. While it would be ideal to propose thresholds for when to update meta-analyses, there
are multiple reasons that simple rules cannot be derived a priori. The decision to update the meta-analyses depends on the relative sample sizes and risk of bias of the new versus prior studies, and the degree to which the new evidence agrees with or contradicts the existing evidence. When few data for a particular outcome or intervention exist, adding new studies may impact the pooled estimates substantially, depending on the details of the new versus the prior evidence, and the SOE of the prior evidence.

2.7 Updating Strength of Evidence Ratings

One of the anticipated challenges was being able to determine the threshold to change the SOE in an LSR, although we did not find this a problem. Following the AHRQ EPC methods guidance on SOE,11 we could determine if new evidence changed the SOE in most cases because the process for determining SOE considers a body of evidence regardless of whether this is a primary review or an update. In our experience, making decisions about the SOE in quarterly updates (to date) has not been difficult. For example, the evidence added in the third quarterly report consisted of observational studies of interventions and outcomes not covered by the previously included studies. The evidence from these new studies was rated insufficient SOE. While we did not encounter this, a theoretical example where the SOE changed would be where new larger trials were added, such that the assessment of the domain of precision moved from imprecise to precise, thus improving the SOE rating.

2.8 Peer Review of Interim Reports

The progress reports were not peer reviewed. Without the external expertise from independent peer reviewers, LSR teams need to include those with a high level of methodological and clinical expertise on the core team. We benefitted by having investigators on the team with experience in both LSRs and subject matter expertise.

2.9 Surveillance Report Frequency

A challenge in conducting LSRs is determining the time between surveillance reports to capture new evidence. In general, it is difficult to predict when new studies will be published and whether they may influence the original conclusions. For our LSR on cannabis and other plant-based compounds, the frequency of quarterly surveillance reports was determined a priori. To date, however, the volume of search results and eligible studies identified following the first progress report has been small and added little evidence of significance, bringing into question whether the additional time and effort invested in frequent surveillance reports is an optimal use of resources.

There are situations, however (such as the SARs-Co-V2 pandemic), when the massive output of publications justifies the need for regular updates of a rapidly evolving literature base.12-16 A future approach may be to allow the newly identified evidence to determine the timing or type of LSR updates. Living reviews published in Annals of Internal Medicine follow such a model17 where the direction of effect and strength of findings from new evidence determines whether the report is a brief update (i.e., new evidence, but not new conclusions), a full systematic review update (i.e., new evidence that changes at least the strength of evidence, possibly the conclusions), or simply a notice of no new evidence. While the resulting reports in this model would likely be useful, this approach poses a challenge to funders, for whom budgetary decisions typically have to be made well in advance.
3. Conclusions

In summary, our experience with LSR methods used in the “Living Systematic Review on Cannabis and Other Plant-Based Treatments for Chronic Pain” identified numerous successes, a few unavoidable limitations, and a few clear ways to improve processes and methods in the first year of this project (Table 2).

Table 2. Summary of LSR methods approaches, challenges, and proposed resolutions

<table>
<thead>
<tr>
<th>Report Process Stage</th>
<th>Changed Planned Methods?</th>
<th>Facilitators of LSR</th>
<th>Barriers to LSR</th>
<th>Proposed Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>No</td>
<td>Ongoing input from Technical Expert Panel will be useful to determining the need and appropriateness of scope changes</td>
<td>Changing inclusion criteria may require revisiting prior inclusion and exclusion decisions or acknowledging that prior decisions were based on different (older) inclusion criteria</td>
<td>Periodically revisit and potentially revise the scope as new/better evidence becomes available or as new formulations/dosing strategies are introduced; ensure that scope changes are clearly described and whether changes are retroactive or going forward</td>
</tr>
<tr>
<td>Electronic Searches</td>
<td>No</td>
<td>Biweekly search result emails allowed more efficient screening of articles</td>
<td>Error in automated email system</td>
<td>Double check that each database is searched as intended</td>
</tr>
<tr>
<td>Study Selection Process</td>
<td>Yes</td>
<td>Ongoing screening was successful, given the small volume of citations</td>
<td>Automation not found helpful, given small and frequent volume of literature</td>
<td>None</td>
</tr>
<tr>
<td>Data Abstraction</td>
<td>Yes</td>
<td>Ongoing screening was successful, given the small volume of literature</td>
<td>Without a prior report to build on, refinement of abstraction processes was needed; learning and working with a new system (SRDR+) was time-consuming</td>
<td>For LSRs conducted for AHRQ, alignment with SRDR+ requires more work; may be another 6 months before processes will be refined</td>
</tr>
<tr>
<td>Risk of Bias</td>
<td>No</td>
<td>Ongoing assessment was successful, given the small volume of literature</td>
<td>None; on-going assessment helped to avoid problems</td>
<td>None</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>No</td>
<td>Updating meta-analysis more frequently could allow for modifications and additional evidence and analyses to be considered</td>
<td>Statistician availability can be variable when not possible to plan in advance; resources required to update meta-analyses must be weighed against usefulness of updating</td>
<td>May need to increase statistician FTE and have available on an ongoing basis, and plan for regular meta-analysis updates before we know if there is new evidence</td>
</tr>
<tr>
<td>Narrative Synthesis</td>
<td>No</td>
<td>None noted</td>
<td>Many issues with how to report new vs. prior evidence, particularly in this first year when there was not a full report to build on yet</td>
<td>Continue to develop format for surveillance reports; get LSR TEP input on format and communication of findings</td>
</tr>
<tr>
<td>SOE</td>
<td>No</td>
<td>None noted, other than updating with new evidence more frequently than in a traditional systematic review</td>
<td>None identified to date; only the third and fourth quarterly progress reports incorporated SOE</td>
<td>None</td>
</tr>
</tbody>
</table>
Millard et al.’s survey of researchers who had participated in an LSR pilot project identified four main issues of LSRs that need special attention. The first was the need to develop methods, procedures, and processes (automated if possible) for ongoing review. We found that we had mixed success in developing truly efficient methods; we did not find automation of article screening to be helpful. However, we found other approaches were efficient, provided there is adequate staffing and funding (e.g., ongoing screening, abstraction, and risk of bias assessment). The second issue identified in the survey was the need to develop rules for when there is enough evidence to signal a change in conclusions. In our LSR experience, we have not yet identified evidence that would alter the SOE for primary outcomes, and hence no changes to conclusions have been made. We anticipate that in the coming year there may be new evidence where a change in conclusions will be triggered by changes in the SOE in the findings. The third area of concern was the need to standardize how new evidence is reflected in frequent publications (i.e., version control). Because the new evidence identified in our most recent LSR updates resulted in insufficient SOE ratings, we have not yet had experience with this issue. We anticipate that the details of displaying the new evidence will require testing and obtaining input from our LSR TEP. Finally, the survey results noted that managing workload consistency and accuracy across time and managing “burnout” was an issue of concern. We have found that more frequent update searches are more manageable, and that dedicated staff with adequate time is important. More time and experience are needed to understand the impact on accuracy or burnout. We suggest that this would be a good area for future evidence synthesis methods investment, studying these and other issues across multiple AHRQ-funded LSRs.


