Methods Research Report

Testing a Machine Learning Tool for Facilitating Living Systematic Reviews of Chronic Pain Treatments

AHRQ

Agency for Healthcare
Research and Quality

Testing a Machine Learning Tool for Facilitating Living Systematic Reviews of Chronic Pain Treatments

Investigators:
Roger Chou, M.D.
Tracy Dana, M.L.S
Kanaka D. Shetty, M.D., M.S.

Key Messages

Purpose of Project

To develop and test text word-only search strategies without MEDLINE® indexing for three chronic pain living reviews, test a machine classifier on studies identified using the text word search strategies, and apply the machine classifier prospectively on a monthly basis to update searches.

Key Messages

- Text word-only searches optimized are associated with high sensitivity but reduced precision compared with standard searches that utilized MeSH indexing terms.
- A machine learning classifier had high recall for identifying studies using text word searches for three systematic reviews of chronic pain; precision was low to moderate.
- Use of the machine learning classifier resulted in a small to moderate estimated time savings when conducting update searches for living systematic reviews.

This report is based on research conducted by the Pacific Northwest Evidence-based Practice Center (EPC) and the Southern California EPC-RAND Corporation under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract Nos. 290-2015-00009-I and 290-2015-00010-I). 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 healthcare 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.

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 healthcare in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new healthcare 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 healthcare system as a whole by providing important information to help improve healthcare quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this Methods Research Project 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.

Gopal Khanna, M.B.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

Stephanie Chang M.D., M.P.H.
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
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:

Justin Clark
Senior Research Information Specialist
Institute for Evidence-Based Healthcare
Faculty of Health Sciences & Medicine
Bond University
Robina, Queensland, Australia

Gerald Gartlehner, M.D., M.P.H.
Associate Director, RTI-UNC Evidence-Based Practice Center
Research Triangle Institute International
Research Triangle Park, NC
Chair and Professor, Department for Evidence-Based Medicine and Clinical Epidemiology
Danube University Krems
Krems, Austria

Testing a Machine Learning Tool for Facilitating Living Systematic Reviews of Chronic Pain Treatments

Structured Abstract

Background. Living systematic reviews can more rapidly and efficiently incorporate new evidence into systematic reviews through ongoing updates. A challenge to conducting living systematic reviews is identifying new articles in a timely manner. Optimizing search strategies to identify new studies before they have undergone indexing in electronic databases and automation using machine learning classifiers may increase the efficiency of identifying relevant new studies.

Methods. This project had three stages: develop optimized search strategies (Stage 1), test machine learning classifier on optimized searches (Stage 2), and test machine learning classifier on monthly update searches (Stage 3). Ovid® MEDLINE® search strategies were developed for three previously conducted chronic pain reviews using standard methods, combining National Library of Medicine Medical Subject Headings (MeSH) terms and text words ("standard searches"). Text word-only search strategies ("optimized searches") were also developed based on the inclusion criteria for each review. In Stage 2, a machine learning classifier was trained and refined using citations from each of the completed pain reviews ("training set") and tested on a subset of more recent citations ("simulated update"), to develop models that could predict the relevant of citations for each topic. In Stage 3, the machine learning models were prospectively applied to "optimized" monthly update searches conducted for the three pain reviews.

Results. In Stage 1, the optimized searches were less precise than the standard searches (i.e., identified more citations that reviewers eventually excluded) but were highly sensitive. In Stage 2, a machine learning classifier using a support vector machine model achieved 96 to 100 percent recall for all topics, with precision of between 1 and 7 percent. Performance was similar using the training data and on the simulated updates. The machine learning classifier excluded 35 to 65 percent of studies classified as low relevance. In Stage 3, the machine classifier achieved 97 to 100 percent sensitivity and excluded (i.e., classified as very low probability) 45 to 76 percent of studies identified in prospective, actual update searches. The estimated savings in time using the machine classifier ranged from 2.0 to 13.2 hours.

Conclusions. Text word-only searches to facilitate the conduct of living systematic reviews are associated with high sensitivity but reduced precision compared with standard searches using MeSH indexing terms. A machine learning classifier had high recall for identifying studies identified using text word searches, but had low to moderate precision, resulting in a small to moderate estimated time savings when applied to update searches.

Contents

Tables

Appendixes

Appendix A. Project Flow Diagram

Appendix B. Original and Text Word-Only Search Strategies

Appendix C. Cases

Background

Systematic reviews become outdated relatively quickly when new evidence becomes available. There is typically a gap of several years between systematic review (SR) updates, which can result in delayed incorporation of new evidence. In addition, inefficiencies may occur in updating, due to changing review personnel, review methods, and processes. Living systematic reviews (LSRs) have been proposed as a method for more rapidly and efficiently incorporating new evidence into systematic reviews. In an LSR, searches are updated frequently on a regular basis, and new evidence is incorporated as it becomes available.[1,2] A key challenge to conducting LSRs is the time and resources required to periodically update the search to identify relevant new articles. Automation using machine learning classifiers is one approach for improving the efficiency of identifying relevant new articles in update searches.[3] A machine learning classifier that accurately classifies citations identified in searches as high or low relevance could save time by reducing the need for investigators to individually review each citation for eligibility.[4] Instead of needing two investigators to perform dual review of each citation for eligibility, for example, integration of a machine learning classifier into a LSR work flow could replace one of the investigators.

The Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) program commissioned a series of three systematic reviews by the Pacific Northwest EPC on chronic pain (*Opioid Treatments for Chronic Pain*, *Nonopioid Pharmacologic Treatments for Chronic Pain*, and *Noninvasive Nonpharmacological Treatments for Chronic Pain: A Systematic Review Update*). Systematic reviews of chronic pain are potential candidates for LSRs due to ongoing interest in the topic and a high volume of new studies. A machine learning classifier is a potential automation tool to facilitate more efficient conduct of LSRs by assessing citations identified in update searches for relevance, potentially reducing the workload on systematic review investigators.[2] AHRQ previously commissioned the Southern California EPC (RAND) to pilot test a previously developed machine learning tool on a systematic review conducted by the Pacific Northwest EPC on *Noninvasive Nonpharmacological Treatment for Chronic Pain* (hereafter referred to as the chronic pain report).[5] Based on the sample of studies identified in searches and included in the final chronic pain report, the machine classifier was associated with a sensitivity of 0.87 and positive predictive value of 0.13 for identifying included studies; the classifier excluded almost 60 percent of irrelevant citations. An observation during pilot testing was that the searches used for the chronic pain report used MeSH term indexing as well as nonindexed search terms; this could limit the usefulness of the searches for LSRs because of the lag in MeSH term indexing of new studies (e.g., for studies in the pre-MEDLINE database). Another observation during pilot testing was that the chronic pain report restricted inclusion to studies with followup of at least 1 month after completing a course of therapy. This information is typically not found in data accessible to the machine learning classifier (i.e., titles and abstracts from the citations); as a result, the machine learning classifier was not able to reliably distinguish between studies included or excluded based on this criterion, therefore limiting its efficiency.

The purpose of this project was to build on and expand upon the prior efforts by developing and testing optimized search strategies without MEDLINE indexing for the three newly commissioned chronic pain reviews (Stage 1), testing the machine classifier on the optimized search strategies using the studies identified in the initial searches to gauge performance using title and abstract information alone (Stage 2), and then applying the machine classifier prospectively on a monthly basis to update searches (Stage 3).

Methods

This project was a collaboration between the Pacific NorthwestEPC, which conducted the initial chronic pain report[5] used by the RAND EPC in earlier testing of the machine learning classifier and conducted the three new chronic pain reports commissioned by AHRQ, and the RAND EPC, which has expertise in developing and testing machine learning classifiers for systematic reviews.[6,7] This project had three stages (Appendix A): develop optimized search strategies (Stage 1), test machine learning classifier on optimized searches (Stage 2), and test machine learning classifier on monthly update searches (Stage 3). The study was conducted according to a predefined protocol developed in consultation with AHRQ; however, the protocol was not preregistered or published.

Stage 1: Develop Optimized Search Strategies

For each chronic pain report, Ovid MEDLINE® search strategies were developed for the three new chronic pain reviews using standard methods, combining National Library of Medicine Medical Subject Headings (MeSH) terms and text words ("standard searches").[8] For the chronic pain reviews that were updates,[5,9] the search strategies were based on the searches conducted for the prior reviews and revised/updated to address any scope changes. Text word-only search strategies ("optimized searches") were also developed based on the population and interventions found in the PICOTS (population, intervention, comparator, outcome, timing, setting) of each systematic review protocol, with further input from principal investigators when clarification was needed on specific interventions (e.g., specific drug names not included in the PICOTS). The complete standard and optimized search strategies are shown in Appendix A. Search strategies were peer reviewed using the validated PRESS (Peer Review of Electronic Search Strategies) instrument.[10] Search strategies were tested against a sample of approximately 25 included randomized controlled trials (cases) for each review, based on dual investigator full-text review for eligibility (Appendixes B and C). The sensitivity (proportion of included citations identified, also known as recall), precision (number of included citations in the search relative to the total number of citations identified), and the number needed to read (NNR, the total number of citations identified divided by the number of included citations in the search),[11] were calculated for both sets of searches for each review.

Stage 2: Test Machine Learning Classifier on Optimized Searches

In Stage 2, we aimed to develop machine learning classifiers using citations from previous reports. We simulated an update by using citations from earlier years ("training data") to develop and refine models that could predict the relevance of more recent citations for each topic. After fine-tuning these topic-specific models, we then simulated LSRs by applying each model to the most recent citations for the topic. We hypothesized that results on these "simulated updates" could be used to estimate performance on true update searches. We describe this approach in greater detail below.

Processing Literature Search Results Into "Gold Standard" Data

After performing the optimized searches for the three chronic pain reviews, we reviewed citations using standard methods (dual investigator review for eligibility at the full-text level) to create a set of "gold standard" results from the searches conducted for the draft report of the three original reports and the initial update searches. To prepare the citations for modeling using a "bag-of-words" approach, we processed the titles and abstracts into a set of word frequencies that were further normalized to de-emphasize words that are common across documents.[12,13] In contrast to prior studies,[7] we did not derive additional metadata from the citation (article length, presence of an abstract, etc.) because they did not appear to contribute substantially to machine classifier performance for these topics. We also did not add metadata based on MeSH indexing terms because such features would be absent from a high proportion of citations obtained in update searches.

Initial Model Training and Testing

For each topic, we developed models that used the processed citations (i.e., the normalized word frequency data and inclusion decisions) from 1960 to 2016 to predict relevance among 2017 citations. Each model used one of the following algorithms to generate predictions: support vector machines (SVM),[14] a multilayer perceptron (a neural network model),[15] regularized logistic regression,[16,17] or synthetic minority over-sampling (SMOTE, which is a resampling algorithm that we used to mitigate the paucity of relevant articles in most systematic review searches).[18] After completing this testing, we opted to use SVM for the remainder of the analyses on each topic because its performance was equivalent to or slightly better than other approaches.

To develop the final SVM models, we set up a process that used standard, open-source software to select optimal model parameters using fivefold cross-validation[16] for each review. Briefly, each training set (in this case, citations published during 1960–2016) was first split into five random subsets. Next, for each potential set of model parameters, a model was derived using four subsets and predictions were generated on the remaining subset; this was repeated five times, one for each "fold." Finally, the final model (with associated parameters) was selected by the software to optimize a combination of recall and precision, as follows. Given the importance of achieving near-perfect recall, we had the process severely penalize models that failed to achieve a minimum of 95 percent recall; among models meeting that criterion, our process then penalized low precision to improve efficiency. These criteria led to selection of models that identified more than 95 percent of relevant articles but subsequently minimized irrelevant studies at the expense of possibly missing small percentages of relevant studies. We then applied the final models to 2017 citations and computed recall, precision, percentage of citations predicted to be included or excluded, and NNR.

Simulating Predictive Performance for Living Systematic Reviews

After deriving final models for each topic as described above, we generated predictions for processed "gold standard" citations from 2018 and early 2019 (which serve as update searches for the purpose of this simulation). We report metrics for evaluating the performance of the machine classifier on each search, including recall, precision, percentage of citations predicted to be included or excluded, and NNR. We then qualitatively compared performance from the initial modeling phase (i.e., 2017 citations) to performance on the simulated update results (i.e.,

citations from 2018 and early 2019) to determine the degree to which performance in the original review predicted performance on subsequent updates.

Stage 3: Test Machine Learning Classifier on Monthly Update Searches

The three machine learning models were prospectively applied to actual update searches conducted for the three chronic pain reviews performed on a monthly basis as the review entered into the draft report and finalization phases. Due to the small number of studies included in each monthly update search, we tested the update searches in two month blocks (July and August 2019 and September and October 2019). The relevance predictions were then compared to final inclusion decisions by Pacific Northwest EPC staff to compute recall, precision, number of citations predicted to be included or excluded, and NNR. Using these metrics, we qualitatively compared the performance on the actual update to that on the simulated update. Results of the machine classifier for studies identified from pre-MEDLINE were reported separately.

The time savings estimation was based on the number of studies classified as excluded by the machine learning classifier (i.e., classified as negligible-probability) multiplied by the average estimated time for an investigator to review each study (0.6 minutes per study [1 hour/100 studies]). Qualitative feedback was also obtained from investigators and staff at the Pacific Northwest EPC on the impact of using the machine learning classifier on work flow.

Results

Stage 1: Develop Optimized Search Strategies

The sensitivity (recall) of the optimized searches was 100 percent (Table 1), or at least as high as the sensitivity of the standard searches (Table 2). As expected, the optimized searches were less efficient than the standard searches (i.e., identified more citations that were eventually excluded by reviewers), with precision ranging from 0.39 to 0.60 percent and a number needed to read (NNR) of 160 to 255, compared with precision of 0.61 to 1.44 percent and NNR of 69 to 163.

Table 1. Standard (MeSH and text word) search strategy testing

Topic	Total Citations Retrieved	Recall (%)	Precision (%)	NNR
Noninvasive, Nonpharmacologic Treatment for Chronic Pain	1,913	90	1.44	69
Nonopioid Pharmacologic Treatments for Chronic Pain	2,008	100	0.61	163
Opioid Treatments for Chronic Pain	2,289	90	0.79	136

NNR = number needed to read

Table 2. Optimized (text word-only) search strategy testing

Topic	Total Citations Retrieved	Recall (%)	Precision (%)	NNR
Noninvasive, Nonpharmacologic Treatment for Chronic Pain	8,704	100	0.60	160
Nonopioid Pharmacologic Treatments for Chronic Pain	4,675	100	0.39	255
Opioid Treatments for Chronic Pain	3,988	100	0.42	236

NNR = number needed to read

Stage 2: Test Machine Learning Classifier on Optimized Searches

After deduplicating citations from prior searches, each study topic had a large set of citations that could be used as training (1960 to 2017) and validation data sets (2018 to 2019), but the data were imbalanced in that <5 percent of citations were included for each topic (Table 3). Citations from the validation data (which were used to simulate updates performed in 2018 and 2019) were more imbalanced (i.e., a lower proportion of studies included) than in the training data.

Table 3. Counts of citations used in Stage 2

Report	Training Data: Total Citations	Training Data: Excluded Full Text	Training Data: Included	Validation Data: Total Citations	Validation Data: Excluded Full Text	Validation Data: Included
Noninvasive, Nonpharmacologic	4,813	4,572	241 (5.0%)	796	781	15 (1.9%)
Nonopioid	7,473	7,232	241 (3.2%)	249	239	10 (4.0%)
Opioid	4,710	4,526	184 (3.9%)	1,444	1,435	9 (0.6%)

In Stage 2, the machine learning classifier models achieved 96 to 100 percent sensitivity for all topics, with precision of between 1 and 7 percent, for a NNR between 15 and 100.

Performance was similar using the training data (2017 citations, Table 4) and on the simulated updates (2018 and 2019 searches, Table 5). Citation exclusion rates varied between 35 and 65 percent.

Table 4. Classifier performance on training data: results on 2017 citations

Report	Total Citations	Citations Included in Report	Citations Predicted by ML Classifier to be Relevant	Citations Excluded by ML Classifier (%)	Recall (%)	Precision (%)	NNR
Noninvasive, Nonpharmacologic	1,407	24	562 (40%)	845 (60%)	96	4	24
Nonopioid	457	14	187 (44%)	256 (56%)	100	7	14
Opioid	1,362	11	1,009 (74%)	353 (26%)	100	1	100

ML = machine learning; NNR = number needed to read

Table 5. Classifier performance on validation data: results on 2018–2019 citations

Report	Total Citations	Citations Included in Report	Citations Predicted by ML Classifier to be Relevant	Citations Excluded by ML Classifier (%)	Recall (%)	Precision (%)	NNR
Noninvasive, Nonpharmacologic	796	15	288 (36%)	508 (64%)	100	5	19
Nonopioid	249	10	143 (57%)	106 (43%)	100	7	14
Opioid	1,444	9	924 (65%)	511 (35%)	100	1	103

NNR = number needed to read

Stage 3: Test Machine Learning Classifier on Monthly Update Searches

When applied to the monthly update searches, the machine learning classifier models achieved 100 percent sensitivity across searches for both the July/August and September/October periods, with the exception of 97 percent sensitivity for the July/August nonpharmacologic therapies search (Table 6). The classifier correctly assigned very low relevance scores to 45 to 76 percent of excluded studies. The machine classifier performance on the updates was consistent with the results on the simulated updates (Table 5), suggesting that simulated updates can be used to establish machine learning performance standards for particular areas. However, the efficiency varied across the specific chronic pain topics and for the first update searches (July/August) versus the second update searches (September/October). For the July/August searches, the precision ranged from 0.5 percent (opioids) to 8 percent (nonpharmacologic therapies). For the September/October searches, precision was lower across reviews, ranging from 0.2 percent (opioids) to 0.9 percent (nonopioids). The search yield (number of citations identified) was higher for the September/October updates compared with the July/August updates, but there were fewer included citations; this was due in part to some included studies being identified during the peer review or public comment process that occurred around the time of the July/August update searches.

Table 6. Classifier performance on update searches

Report	Total Citations	Citations Included in Report	Citations Predicted by ML Classifier to be Relevant (%)	Citations Excluded by ML Classifier (%)	Recall (%)	Precision (%)	NNR
Noninvasive, Nonpharmacologic • July/August	1,536	30	375 (24%)	1161 (76%)	97	8	13
• Sept/Oct	1,755	3	433 (25%)	1322 (75%)	100	0.7	144
Nonopioid • July/August	345	7	148 (43%)	197 (57%)	100	5	21
• Sept/Oct	418	1	113 (27%)	305 (73%)	100	0.9	113
Opioid • July/August	763	2	418 (55%)	345 (45%)	100	0.5	209
• Sept/Oct	1,097	1	580 (53%)	517 (47%)	100	0.2	580

ML = machine learning; NNR = number needed to read

The machine learning classifier excluded (i.e., classified as very low probability) 45 to 76 percent of studies in the two update searches. The number of studies excluded by the machine learning classifier ranged from 197 (nonopioid medication therapies review, July/August) to 1,322 (nonpharmacologic therapies review, September/October). Assuming that one investigator would not need to assess studies excluded by the machine learning classifier, the estimated savings in time ranged from 2.0 to 13.2 hours.

EPC staff indicated a modest time investment to develop efficient processes for providing data from the update searches to the investigator running the machine classifier and to receive and interpret the output, but once these processes were in place, reported that utilizing the machine learning classifier had minimal impact on work flow.

Discussion

Our study focused on methods for identifying studies for inclusion in LSRs. It had two main objectives: to develop and test optimized search strategies without MEDLINE indexing and to test the performance of a machine classifier on the optimized search strategies. The searches were based on those conducted for three AHRQ-commissioned, complementary reviews on chronic pain.

Our results indicate that optimized text word-only searches can be developed with high sensitivity. The use of text word-only searches is important for conducting LSRs because of the lag between entry of studies into the Pre-MEDLINE database and tagging with MeSH term indexing. Therefore, relying on index terms in the searches would result in delayed identification of new studies. The optimized text word searches had sensitivity at least as high as standard searches that utilized MeSH terms. However, the optimized searches were associated with reduced efficiency—that is, a higher NNR, an expected tradeoff since they relied on text words.

We also found that the machine learning classifier had 100 percent recall in simulated updates using a validation set of studies from the initial searches, as well as actual monthly update searches. This suggests that researchers may be able to estimate performance of machine learning classifiers using simulated updates. Precision ranged from moderate (NNR 13) to low (NNR 580). The variability in precision could be due in part to the breadth of the reviews: precision was lowest for a review on opioids for chronic pain that included questions on predictive utility, observational studies, and a wide range of interventions and comparisons, including various opioids, opioid dosing strategies, and risk mitigation strategies. By contrast, precision was higher for reviews on nonpharmacologic therapies and nonopioid pharmacologic therapies for chronic pain that focused on randomized trials and did not include questions on dosing, risk mitigation strategies, or diagnostic accuracy. Precision was probably limited by the models' reliance on title and abstract information, which may not provide sufficient information for determining inclusion. For example, the review on noninvasive, nonpharmacologic therapies was restricted to studies with at least 1-month followup after completion of therapy, but the abstract often only reports the duration of therapy, or the total duration of follow-up (including the duration of therapy). It is possible that precision could be improved if the machine learning classifier utilized the full text of the article.

The estimated time savings ranged from small to moderate, totaling 2.0 to 13.2 hours when update searches were combined in two month blocks (1.0 to 6.1 hours per month). Even though the machine learning classifier was least precise for the opioids review, the time savings was moderate, given the relatively large number of citations identified in the update searches. Incorporating the machine learning classifier into the update search process appeared to have minimal impact on work flow once initial processes were set up, suggesting that little of the time savings from the reduced number of citations requiring screening would be offset by more inefficient work flow.

Our study had limitations. The sample was three reviews in a specific topic area (chronic pain). The applicability to other reviews and topics is uncertain. A sample of approximately 25 studies was used to test the optimized search strategies for each review, based on the inclusion decisions available at the time that the searches were developed. The optimal number of studies to test optimized search strategies is not known; however, the applicability of our approach to small reviews with few included studies may be limited. Time savings were estimated based on the average time required to review a title/abstract. However, the time savings is likely to vary depending on the experience of the reviewers, the complexity of the review, the types of study

8

designs included, familiarity with the topic, and other factors. In addition, the time savings is likely to vary for different citations; the time savings is likely to be less if the citations categorized by the machine learning classifier as irrelevant are easy to exclude by an investigator. Another limitation is that the machine learning classifier was only tested on two monthly update searches with a small number of included citations, particularly for the second update search. The update searches were imbalanced (i.e., small number of included citations relative to the number of excluded citations), particularly for the second set of update searches. Therefore, sensitivity estimates for the update searches are not robust. In addition, standard machine learning algorithms have difficulty with imbalanced data because predicting irrelevance for all studies would be more than 98 percent accurate but completely unhelpful. More data are needed to confirm the machine learning classifier's performance on update searches. Because the machine learning classifier missed only one study in the simulated updates, it was not possible to determine how missed studies could have impacted the conclusions of the LSR. Model development prioritized very high sensitivity, reducing the efficiency of the machine learning classifier. Training the models using lower recall thresholds could result in different performance, but we did not test such approaches. Finally, this project tested multiple algorithms using software that was customized for this task; applying this machine learning classifier to other projects would require 1−3 days of additional effort to process the citations into a suitable format and set up the classifier. However, the underlying algorithms that underwent testing (SVM, multilayer perceptron, etc.) are open-source, freely available, and could easily be adapted for use by other researchers.[16] Testing of multiple algorithms could be accomplished in an automated fashion (i.e., without programmer input) but would require additional programming effort up front. Alternatively, if citation management software systems added "off-the-shelf" machine learning classifiers to assist with updates, this would likely eliminate the need for additional programming effort by users while retaining most of the time-saving benefits of machine learning.

Conclusion

Text word-only searches to facilitate the conduct of LSRs are feasible and associated with high sensitivity but reduced efficiency compared with standard searches that utilized MeSH indexing terms. A machine learning classifier had high recall for identifying studies identified using text word searches for three systematic reviews of chronic pain, but had low to moderate precision, resulting in a small to moderate estimated time savings when conducting update searches.
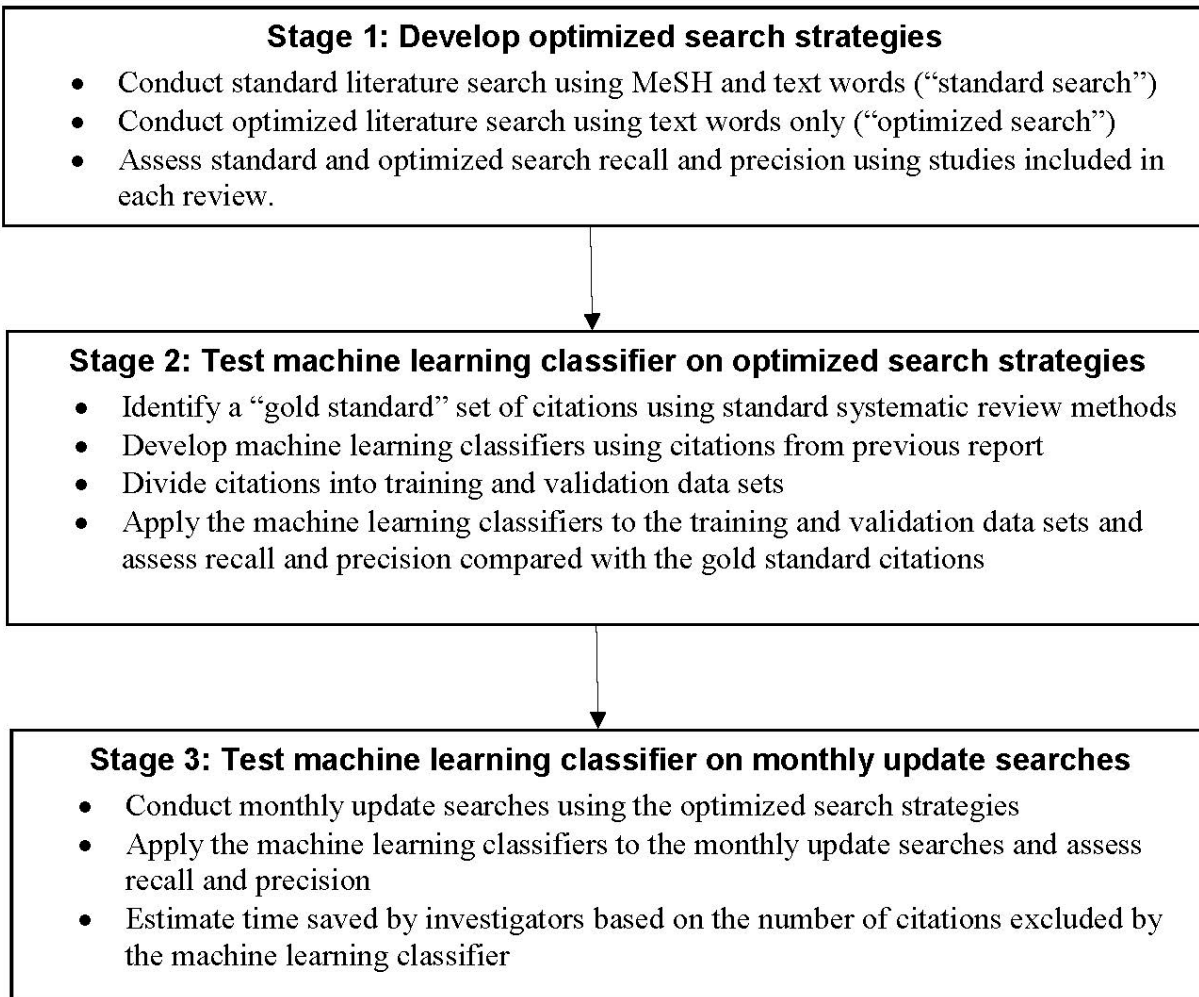
9

References

1   Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017 Nov;91:23-30. doi: 10.1016/j.jclinepi.2017.08.010. PMID: 28912002.

2   Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014 Feb;11(2):e1001603. doi: 10.1371/journal.pmed.1001603. PMID: 24558353.

3   Hemens BJ, Iorio A. Computer-Aided Systematic Review Screening Comes of Age. Ann Intern Med. 2017 Aug 1;167(3):210-1. doi: 10.7326/m17-1295. PMID: 28605802.

4   Thomas J, Noel-Storr A, Marshall I, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017 Nov;91:31-7. doi: 10.1016/j.jclinepi.2017.08.011. PMID: 28912003.

5   Skelly AC, Chou R, Dettori JR, et al. AHRQ Comparative Effectiveness Reviews. Noninvasive Nonpharmacological Treatment for Chronic Pain: A Systematic Review. Rockville (MD): Agency for Healthcare Research and Quality (US); 2018.

6   Dalal SR, Shekelle PG, Hempel S, et al. A pilot study using machine learning and domain knowledge to facilitate comparative effectiveness review updating. Med Decis Making. 2013 Apr;33(3):343-55. doi: 10.1177/0272989x12457243. PMID: 22961102.

7   Shekelle PG, Shetty K, Newberry S, et al. Machine Learning Versus Standard Techniques for Updating Searches for Systematic Reviews: A Diagnostic Accuracy Study. Ann Intern Med. 2017 Aug 1;167(3):213-5. doi: 10.7326/l17-0124. PMID: 28605762.

8   Cooper C, Booth A, Varley-Campbell J, et al. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Med Res Methodol. 2018 Aug 14;18(1):85. doi: 10.1186/s12874-018-0545-3. PMID: 30107788.

9   Chou R, Deyo R, Devine B, et al. The Effectiveness and Risks of Long-Term Opioid Treatment of Chronic Pain. Evid Rep Technol Assess (Full Rep). 2014 Sep(218):1-219. doi: 10.23970/ahrqepcerta218. PMID: 30313000.

10 McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol. 2016 Jul;75:40-6. doi: 10.1016/j.jclinepi.2016.01.021. PMID: 27005575.

11 Golder S, Mason A, Spilsbury K. Systematic searches for the effectiveness of respite care. Journal of the Medical Library Association. 2008 Apr;96(2):147-52. doi: https://dx.doi.org/10.3163/1536-5050.96.2.147. PMID: 18379671.

12 Ramos J. Using tf-idf to determine word relevance in document queries. Proceedings of the first instructional conference on machine learning. 2003;242:133-42.

13 Jones K. IDF term weighting and IR research lessons. Journal of Documentation - J DOC. 2004 10/01;60:521-3. doi: 10.1108/00220410410560591.

14 Nédellec C, Rouveirol C, eds. Text categorization with Support Vector Machines: Learning with many relevant features. Machine Learning: ECML-98; 1998 1998//; Berlin, Heidelberg. Springer Berlin Heidelberg.

15 Pal SK, Mitra S. Multilayer perceptron, fuzzy sets, and classification. IEEE Transactions on Neural Networks. 1992;3(5):683-97. doi: 10.1109/72.159058.

16 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011;12:2825-30.

17 Zhang T. Solving large scale linear prediction problems using stochastic gradient descent algorithms; 2004.

18 Chawla N, Bowyer K, Hall L, et al. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. (JAIR). 2002 01/01;16:321-57. doi: 10.1613/jair.953.

Abbreviations and Acronyms

AHRQ	Agency for Healthcare Research and Quality
EPC	Evidence-based Practice Center
LSR	living systematic review
NNR	number needed to read
SMOTE	Synthetic Minority Oversampling Technique
SR	systematic review
SVM	support vector machine

Appendix A. Project Flow Diagram

Stage 1: Develop optimized search strategies

- Conduct standard literature search using MeSH and text words ("standard search")
- Conduct optimized literature search using text words only ("optimized search")
- Assess standard and optimized search recall and precision using studies included in each review.

Stage 2: Test machine learning classifier on optimized search strategies

- Identify a "gold standard" set of citations using standard systematic review methods
- Develop machine learning classifiers using citations from previous report
- Divide citations into training and validation data sets
- Apply the machine learning classifiers to the training and validation data sets and assess recall and precision compared with the gold standard citations

Stage 3: Test machine learning classifier on monthly update searches

- Conduct monthly update searches using the optimized search strategies
- Apply the machine learning classifiers to the monthly update searches and assess recall and precision
- Estimate time saved by investigators based on the number of citations excluded by the machine learning classifier

Appendix B. Original and Text Word-Only Search Strategies

Original (MeSH and text word) search strategies
Noninvasive Nonpharmacologic Treatment for Chronic Pain

1. exp Low Back Pain/ or ((back or spine or spinal) adj2 pain).ti,ab.
2. exp Chronic Pain/
3. Neck Pain/ or neck.ti,ab.
4. exp Osteoarthritis/ or osteoarthritis.ti,ab.
5. Headache/ or headache.ti,ab.
6. Fibromyalgia/ or fibromyalgia.ti,ab.
7. exp Exercise Therapy/
8. exp Physical Therapy Modalities/
9. exp Braces/
10. exp Mind-Body Therapies/
11. exp Acupuncture Therapy/
12. exp Rehabilitation/
13. exp Psychotherapy/
14. exp Musculoskeletal Manipulations/
15. (noninvasive or non-invasive or nonpharmacologic* or non-pharmacologic*).ti,ab.
16. (exercise or physical therapy or cognitive or behavioral or feedback or relaxation or acceptance or commitment or traction or ultrasound or stimulation or laser or magnet* or inferential or electromuscular or diathermy or heat or cold or manipulation or manual or craniosacral or mindfulness or meditation or mind-body or yoga or pilates or Qigong or acupuncture or functional restoration or multidisciplin* or interdisciplin*).ti,ab.
17. rh.fs.
18. or/1-6
19. or/7-17
20. 18 and 19
21. randomized controlled trial.pt.
22. controlled clinical trial.pt.
23. clinical trials as topic.sh.
24. (random* or trial or placebo).ti,ab.
25. clinical trials as topic.sh.
26. exp animals/ not humans.sh.
27. or/21-25
28. 27 not 26
29. 20 and 28
30. limit 20 randomized controlled trial
31. 29 or 30
32. limit 31 to yr=2015-current

Nonopioid Pharmacologic Treatments for Chronic Pain

1. (acetaminophen or celecoxib or diclofenac or diflunisal or etodolac or fenoprofen or flurbiprofen or ibuprofen or indomethacin or ketoprofen or ketorolac or meclofenamate or "mefenamic acid" or meloxicam or nabumetone or naproxen or oxaprozin or paracetamol or piroxicam or salsalate or sulindac or tenoxicam or "tiaprofenic acid" or tolmetin).ab,kw,sh,ti.
2. (carbamazepine or gabapentin or oxcarbazepine or pregabalin).ab,kw,sh,ti.
3. (desvenlafaxine or duloxetine or levomilnacipran or milnacipran or venlafaxine).ab,kw,sh,ti.
4. (amitriptyline or desipramine or doxepin or imipramine or nortriptyline).ab,kw,sh,ti.
5. (capsaicin or cannabis or marijuana or cannabinoid* or memantine).ab,kw,sh,ti.
6. (topical adj2 lidocaine).ab,kw,ti.
7. or/1-6
8. exp Neuralgia/
9. Fibromyalgia/
10. exp Anemia, Sickle Cell/
11. Headache/
12. exp Headache Disorders/
13. Musculoskeletal Pain/
14. exp Osteoarthritis/
15. Low Back Pain/
16. Neck Pain/
17. exp Arthritis, Rheumatoid/
18. Spondylitis, Ankylosing/
19. ("ankylosing spondylitis" or "neuropathic pain" or neuropathy or fibromyalgia or "sickle cell" or headache or "musculoskeletal pain" or osteoarthritis or "low back pain" or "neck pain" or "inflammatory pain" or "rheumatoid arthritis").ab,kw,ti.
20. or/8-19
21. 7 and 20
22. randomized controlled trial.pt.
23. controlled clinical trial.pt.
24. clinical trials as topic.sh.
25. (random* or trial or placebo).ti,ab.
26. clinical trials as topic.sh.
27. exp animals/ not humans.sh.
28. or/22-26
29. 28 not 27
30. 21 and 29
31. limit 21 to randomized controlled trial
32. 30 or 31
33. limit 32 to yr=2009-current

Opioid Treatments for Chronic Pain
KQ 1-3 (Treatment)
  1. Chronic Pain/
  2. exp arthralgia/ or exp back pain/ or cancer pain/ or exp headache/ or exp musculoskeletal pain/ or neck pain/ or exp neuralgia/ or exp nociceptive pain/ or pain, intractable/ or fibromyalgia/ or myalgia/

3. Pain/
4. chronic.ti,ab,kw.
5. 3 and 4
6. ((chronic or persistent or intractable or refractory) adj1 pain).ti,ab,kw.
7. ((((back or spine or spinal or leg or musculoskeletal or neuropathic or nociceptive or radicular) adj1 pain) or headache or arthritis or fibromyalgia or neuropathy or osteoarthritis).ti,ab,kw.
8. 1 or 2 or 5 or 6 or 7
9. exp Analgesics, Opioid/
10. opioid*.ti,ab,kw.
11. (buprenorphine or codeine or fentanyl or hydrocodone or hydromorphone or methadone or morphine or oxycodone or oxymorphone or tapentadol).ti,ab,kw,sh,hw.
12. 9 or 10 or 11
13. 8 and 12
14. 13 not (intravenous or intramuscular or injection* or intrathecal or epidural or block or preoperative or perioperative or acute).ti.
15. randomized controlled trial.pt.
16. controlled clinical trial.pt.
17. clinical trials as topic.sh.
18. (random* or trial or placebo).ti,ab.
19. clinical trials as topic.sh.
20. exp animals/ not humans.sh.
21. or/15-19
22. 21 not 20
23. 14 and 22
24. limit 14 to randomized controlled trial
25. 23 or 24
26. limit 25 to yr=2009-current

Text word-only search strategies
Noninvasive Nonpharmacologic Treatment for Chronic Pain
1. ((chronic or pain) and (back or spine or spinal or cervical or radicular or neck or knee or hip)).ti,ab,kw
2. (osteoarthritis or headache or fibromyalgia).ti,ab,kw
3. (noninvasive or non-invasive or nonpharmacologic* or non-pharmacologic*).ti,ab,kw
4. (exercise or "physical therap*").ti,ab,kw
5. (psychotherap* or psycholog* or cognitive or behavior* or CBT or biofeedback or feedback or relaxation or acceptance or commitment).ti,ab,kw
6. (traction or ultrasound or transcutaneous electric* or TENS or laser or interferential or electromuscular or diathermy or heat or cold or brace* or bracing or magnet*).ti,ab,kw
7. (manipulation or manual or massage or craniosacral).ti,ab,kw
8. (mindfulness or meditation or meditative or mind-body or yoga or "tai chi" or "tai ji" or qigong).ti,ab,kw
9. (acupuncture or acupressure).ti,ab,kw
10. ("functional restoration" or kinesiology).ti,ab,kw
11. 1 or 2

12. 11 not (postoperative or post operative).ti,ab,kw
13. or/3-10
14. 12 and 13
15. 14 and (random* or control* or placebo or sham or trial).ti,ab,kw

Nonopioid Pharmacologic Treatments for Chronic Pain
1. ((chronic or pain) and (back or spine or spinal or cervical or radicular or neck or knee or hip)).ti,ab,kw.
2. (chronic adj2 pain).ti,ab,kw.
3. ("ankylosing spondylitis" or "neuropathic pain" or neuropathy or fibromyalgia or "sickle cell" or headache or "musculoskeletal pain" or osteoarthritis or "low back pain" or "neck pain" or "inflammatory pain" or "rheumatoid arthritis").ti,ab,kw.
4. or/1-3
5. acetaminophen.ti,ab,kw.
6. (celecoxib or diclofenac or diflunisal or etodolac or fenoprofen or flurbiprofen or ibuprofen or indomethacin or ketoprofen or ketorolac or meclofenamate or "mefenamic acid" or meloxicam or nabumetone or naproxen or oxaprozin or piroxicam or salsalate or sulindac or tenoxicam or "tiaprofenic acid" or tolmetin).ti,ab,kw.
7. (desvenlafaxine or duloxetine or levomilnacipran or milnacipran or venlafaxine).ti,ab,kw.
8. (amitriptyline or desipramine or doxepin or imipramine or nortriptyline).ti,ab,kw.
9. (carbamazepine or gabapentin or oxcarbazepine or pregabalin).ti,ab,kw.
10. (alprazolam or chlordiazepoxide or clobazam or clonazepam or clorazepate or diazepam or estazolam or flurazepam or lorazepam or oxazepam or temazepam or triazolam).ti,ab,kw.
11. (capsaicin or memantine or cannabis or marijuana or cannabidiol or phytocannabinoid* or dronabinol or nabilone).ti,ab,kw.
12. (topical adj2 lidocaine).ti,ab,kw.
13. (alprazolam or chlordiazepoxide or clobazam or clonazepam or clorazepate or diazepam or estazolam or flurazepam or lorazepam or oxazepam or temazepam or triazolam).ti,ab,kw.
14. or/5-13
15. 4 and 14
16. 15 and (random* or control* or placebo or sham or trial).ti,ab,kw.

Opioid Treatments for Chronic Pain
KQ 1-3 (Treatment)
1. ((chronic or pain) and (back or spine or spinal or cervical or radicular or neck or knee or hip)).ti,ab,kw.
2. (chronic adj2 pain).ti,ab,kw.
3. ("ankylosing spondylitis" or "neuropathic pain" or neuropathy or fibromyalgia or "sickle cell" or headache or "musculoskeletal pain" or osteoarthritis or "low back pain" or "neck pain" or "inflammatory pain" or "rheumatoid arthritis").ti,ab,kw.
4. or/1-3
5. opioid*.ti,ab,kw.
6. (buprenorphine or codeine or fentanyl or hydrocodone or hydromorphone or methadone or morphine or oxycodone or oxymorphone or tapentadol or tramadol).ti,ab,kw.
7. 5 or 6
8. 4 and 7

9. 8 not (intravenous or intramuscular or injection* or intrathecal or epidural or block or preoperative or perioperative or acute).ti.
10. 9 and (random* or control* or placebo or sham or trial).ti,ab,kw

Appendix C. Cases

Noninvasive Nonpharmacologic Treatment for Chronic Pain

Banth S, Ardebil MD. Effectiveness of mindfulness meditation on pain and quality of life of patients with chronic low back pain. Int J Yoga. 2015;8(2):128-133. PMID: 26170592.

Bennell KL, Ahamed Y, Jull G, et al. Physical therapist-delivered pain coping skills training and exercise for knee osteoarthritis: randomized controlled trial. Arthritis Care Res (Hoboken). 2016;68(5):590-602. PMID: 26417720.

Blodt S, Pach D, Kaster T, et al. Qigong versus exercise therapy for chronic low back pain in adults--a randomized controlled non-inferiority trial. European Journal of Pain. 2015;19(1):123-131. PMID: 24902673.

Bono F, Salvino D, Mazza MR, et al. The influence of ictal cutaneous allodynia on the response to occipital transcutaneous electrical stimulation in chronic migraine and chronic tension-type headache: a randomized, sham-controlled study. Cephalalgia. 2015;35(5):389-398. PMID: 25078717.

Cash E, Salmon P, Weissbecker I, et al. Mindfulness meditation alleviates fibromyalgia symptoms in women: results of a randomized clinical trial. Ann Behav Med. 2015;49(3):319-330. PMID: 25425224.

Castel A, Castro S, Fontova R, et al. Body mass index and response to a multidisciplinary treatment of fibromyalgia. Rheumatol Int. 2015;35(2):303-314. PMID: 25080875.

Cherkin DC, Anderson ML, Sherman KJ, et al. Two-year follow-up of a randomized clinical trial of mindfulness-based stress reduction vs cognitive behavioral therapy or usual care for chronic low back pain. JAMA. 2017;317(6):642-644.

Cherkin DC, Sherman KJ, Balderson BH, et al. Effect of Mindfulness-Based Stress Reduction vs Cognitive Behavioral Therapy or Usual Care on Back Pain and Functional Limitations in Adults With Chronic Low Back Pain: A Randomized Clinical Trial. JAMA. 2016;315(12):1240-1249. PMID: 28196244.

Groessl EJ, Liu L, Chang DG, et al. Yoga for military veterans with chronic low back pain: a randomized clinical trial. Am J Prev Med. 2017;53(5):599-608. PMID: 28735778.

Helminen EE, Sinikallio SH, Valjakka AL, et al. Effectiveness of a cognitive-behavioural group intervention for knee osteoarthritis pain: a randomized controlled trial. Clin Rehabil. 2015;29(9):868-881. PMID: 25413168.

Herman PM, Anderson ML, Sherman KJ, et al. Cost-effectiveness of mindfulness-based stress reduction versus cognitive behavioral therapy or usual care among adults with chronic low back pain. Spine. 2017;42(20):1511-1520. PMID: 28742756.

Ho LF, Lin ZX, Leung AWN, et al. Efficacy of abdominal acupuncture for neck pain: A randomized controlled trial. PLoS ONE. 2017;12(7):e0181360. PMID: 28715459.

Larsson A, Palstam A, Lofgren M, et al. Resistance exercise improves muscle strength, health status and pain intensity in fibromyalgia--a randomized controlled trial. Arthritis Res Ther. 2015;17:161. PMID: 26084281.

Lauche R, Stumpe C, Fehr J, et al. The effects of Tai Chi and neck exercises in the treatment of chronic nonspecific neck pain: a randomized controlled trial. J Pain. 2016;17(9):1013-1027. PMID: 27345663.

Li X, Lin C, Liu C, et al. Comparison of the effectiveness of resistance training in women with chronic computer-related neck pain: a randomized controlled study. Int Arch Occup Environ Health. 2017;90(7):673-683. PMID: 28528354.

MacPherson H, Tilbrook H, Richmond S, et al. Alexander Technique lessons or acupuncture sessions for persons with chronic neck pain: a randomized trial. Ann Intern Med. 2015;163(9):653-662. PMID: 26926698.

Morone NE, Greco CM, Moore CG, et al. A mind-body program for older adults with chronic low back pain: a randomized clinical trial. JAMA Intern Med. 2016;176(3):329-337. PMID: 26903081.

Natour J, Cazotti Lde A, Ribeiro LH, et al. Pilates improves pain, function and quality of life in patients with chronic low back pain: a randomized controlled trial. Clin Rehabil. 2015;29(1):59-68. PMID: 24965957.

Paolucci T, Piccinini G, Iosa M, et al. Efficacy of extremely low-frequency magnetic field in fibromyalgia pain: A pilot study. J Rehabil Res Dev. 2016;53(6):1023-1034. PMID: 28475205.

Paolucci T, Vetrano M, Zangrando F, et al. MMPI-2 profiles and illness perception in fibromyalgia syndrome: The role of therapeutic exercise as adapted physical activity. J Back Musculoskeletal Rehabil. 2015;28(1):101-109. PMID: 25061029.

Salvat I, Zaldivar P, Monterde S, et al. Functional status, physical activity level, and exercise regularity in patients with fibromyalgia after multidisciplinary treatment: retrospective analysis of a randomized controlled trial. Rheumatol Int. 2017;37(3):377-387. PMID: 27844124.

Sanudo B, Carrasco L, de Hoyo M, et al. Vagal modulation and symptomatology following a 6-month aerobic exercise program for women with fibromyalgia. Clin Exp Rheumatol. 2015;33(1 Suppl 88):S41-45. PMID: 25786042.

Saper RB, Lemaster C, Delitto A, et al. Yoga, physical therapy, or efor chronic low back pain: a randomized noninferiority trial. Ann Intern Med. 2017;167(2):85-94. PMID: 28631003.

Saral I, Sindel D, Esmaeilzadeh S, et al. The effects of long- and short-term interdisciplinary treatment approaches in women with fibromyalgia: a randomized controlled trial. Rheumatol Int. 2016;36(10):1379-1389. PMID: 27055444.

Seferiadis A, Ohlin P, Billhult A, Gunnarsson R. Basic body awareness therapy or exercise therapy for the treatment of chronic whiplash associated disorders: a randomized comparative clinical trial. Disabil Rehabil. 2016;38(5):442-451. PMID: 25955823.

Segal NA, Glass NA, Teran-Yengle P, et al. Intensive gait training for older adults with symptomatic knee osteoarthritis. Am J Phys Med Rehabil. 2015;94(10 Suppl 1):848-858.

Teirlinck CH, Luijsterburg PA, Dekker J, et al. Effectiveness of exercise therapy added to general practitioner care in patients with hip osteoarthritis: a pragmatic randomized controlled trial. Osteoarthritis Cartilage. 2016;24(1):82-90. PMID: 26254237.

Vas J, Santos-Rey K, Navarro-Pablo R, et al. Acupuncture for fibromyalgia in primary care: a randomised controlled trial. Acupunct Med. 2016;34(4):257-266. PMID: 26879181.

Yildiz SK, Ozkan FU, Aktas I, et al. The effectiveness of ultrasound treatment for the management of knee osteoarthritis: a randomized, placebo-controlled, double-blind study. Turk J Med Sci. 2015;45(6):1187-1191. PMID: 26775369.

Zgierska AE, Burzinski CA, Cox J, et al. Mindfulness meditation and cognitive behavioral therapy intervention reduces pain severity and sensitivity in opioid-treated chronic low back pain: findings from a randomized controlled trial. Pain Medicine. 2016;17(10):1865-1881.

Nonopioid Pharmacologic Treatments for Chronic Pain

Allen R, Sharma U, Barlas S. Clinical experience with desvenlafaxine in treatment of patients with fibromyalgia syndrome. Clin Pharmacol Drug Dev. 2017;6(3):224-233. PMID: 27139158.

Atkinson JH, Slater MA, Capparelli EV, et al. A randomized controlled trial of gabapentin for chronic low back pain with and without a radiating component. Pain. 2016;157(7):1499-1507. PMID: 26963844.

Barthel HR, Haselwood D, Longley S, 3rd, Gold MS, Altman RD. Randomized controlled trial of diclofenac sodium gel in knee osteoarthritis. Semin Arthritis Rheum. 2009;39(3):203-212. PMID: 19932833.

Chappell AS, Littlejohn G, Kajdasz DK, et al. A 1-year safety and efficacy study of duloxetine in patients with fibromyalgia. Clin J Pain. 2009;25(5):365-375. PMID: 19454869.

Devi P, Madhu K, Ganapathy B, et al. Evaluation of efficacy and safety of gabapentin, duloxetine, and pregabalin in patients with painful diabetic peripheral neuropathy. Indian J Pharmacol. 2012;44(1):51-56. PMID: 22345870.

Irving G, Tanenberg RJ, Raskin J, et al. Comparative safety and tolerability of duloxetine vs. pregabalin vs. duloxetine plus gabapentin in patients with diabetic peripheral neuropathic pain. Int J Clin Pract. 2014;68(9):1130-1140. PMID: 24837444.

Mahmood R, Jawed I, Khan MI, et al. Comparative role of pregabalin and carbamazepine regarding efficacy in painful diabetic neuropathy. Pak J Pharm Sci. 2017;30(4):1275-1278. PMID: 29039325.

Mathieson S, Maher CG, McLachlan AJ, et al. Trial of pregabalin for acute and chronic sciatica. N Engl J Med. 2017;376(12):1111-1120. PMID: 28328324.

Rauck R, Makumi CW, Schwartz S, et al. A randomized, controlled trial of gabapentin enacarbil in subjects with neuropathic pain associated with diabetic peripheral neuropathy. Pain Practice. 2013;13(6):485-496. PMID: 23186035.

Schukro RP, Oehmke MJ, Geroldinger A, et al. Efficacy of duloxetine in chronic low back pain with a neuropathic component: a randomized, double-blind, placebo-controlled crossover trial. Anesthesiology. 2016;124(1):150-158. PMID: 26517858.

Simon LS, Grierson LM, Naseer Z, Bookman AA, Zev Shainhouse J. Efficacy and safety of topical diclofenac containing dimethyl sulfoxide (DMSO) compared with those of topical placebo, DMSO vehicle and oral diclofenac for knee osteoarthritis. Pain. 2009;143(3):238-245. PMID: 19380203.

Tanenberg RJ, Clemow DB, Giaconia JM, et al. Duloxetine compared with pregabalin for diabetic peripheral neuropathic pain management in patients with suboptimal pain response to gabapentin and treated with or without antidepressants: a post hoc analysis. Pain Pract. 2014;14(7):640-648. PMID: 24152185.

Uchio Y, Enomoto H, Alev L, et al. A randomized, double-blind, placebo-controlled Phase III trial of duloxetine in Japanese patients with knee pain due to osteoarthritis. J Pain Res. 2018;11:809-821. PMID: 29713194.

Urquhart DM, Wluka AE, van Tulder M, et al. Efficacy of low-dose amitriptyline for chronic low back pain: a randomized clinical trial. JAMA Intern Med. 2018;178(11):1474-1481. PMID: 30285054.

van de Donk T, Niesters M, Kowal MA, et al. An experimental randomized study on the analgesic effects of pharmaceutical-grade cannabis in chronic pain patients with fibromyalgia. Pain. 2019;160(4):860-869. PMID: 30585986.

Opioid Treatments for Chronic Pain

Afilalo M, Etropolski MS, Kuperwasser B, et al. Efficacy and safety of Tapentadol extended release compared with oxycodone controlled release for the management of moderate to severe chronic pain related to osteoarthritis of the knee: a randomized, double-blind, placebo- and

active-controlled phase III study. Clinical Drug Investigation. 2010;30(8):489-505. PMID: 20586515.

Baron R, Martin-Mola E, Muller Met al. Effectiveness and safety of Tapentadol Prolonged Release (PR) versus a combination of Tapentadol PR and Pregabalin for the management of severe, chronic low back pain with a neuropathic component: a randomized, d-blind, phase 3b study. Pain Practice. 2015;15(5):455-470. PMID: 24738609.

Breivik H, Ljosaa TM, Stengaard-Pedersen K, et al. A 6-months, randomised, placebo-controlled evaluation of efficacy and tolerability of a low-dose 7-day buprenorphine transdermal patch in osteoarthritis patients naive to potent opioids. Scand J Pain. 2010;1(3):122-141. PMID: 29913983.

Cloutier C, Taliano J, O'Mahony W, et al. Controlled-release oxycodone and naloxone in the treatment of chronic low back pain: a placebo-controlled, randomized study. Pain Res Manag. 2013;18(2):75-82. PMID: 23662289.

DeLemos BP, Xiang J, Benson C, et al. Tramadol hydrochloride extended-release once-daily in the treatment of osteoarthritis of the knee and/or hip: a double-blind, randomized, dose-ranging trial. Am J Ther. 2011;18(3):216-226. PMID: 20215961.

Gimbel J, Spierings EL, Katz N, Xiang Q, Tzanis E, Finn A. Efficacy and tolerability of buccal buprenorphine in opioid-experienced patients with moderate to severe chronic low back pain: results of a phase 3, enriched enrollment, randomized withdrawal study. Pain. 2016;157(11):2517-2526. PMID: 27434505.

Hale ME, Zimmerman TR, Eyal E, et al. Efficacy and safety of a hydrocodone extended-release tablet formulated with abuse-deterrence technology in patients with moderate-to-severe chronic low back pain. J Opioid Manag. 2015;11(6):507-518. PMID: 26728648.

Katz N, Kopecky EA, O'Connor M, Brown RH, Fleming AB. A phase 3, multicenter, randomized, double-blind, placebo-controlled, safety, tolerability, and efficacy study of Xtampza ER in patients with moderate-to-severe chronic low back pain. Pain. 2015;156(12):2458-2467. PMID: 26262828.

Krebs EE, Gravely A, Nugent S, et al. Effect of opioid vs nonopioid medications on pain-related function in [atients with chronic back pain or hip or knee osteoarthritis pain: The SPACE randomized clinical trial. JAMA. 2018;319(9):872-882. PMID: 29509867.

Lange B, von Zabern D, Elling C, Dubois C. Efficacy and safety of tapentadol prolonged release for moderate-to-severe chronic osteoarthritis knee pain: a pooled analysis of two double-blind, randomized, placebo- and oxycodone controlled release-controlled studies. Curr Med Res Opin. 2017;33(8):1413-1422. PMID: 30200781.

Rauck R, Rapoport R, Thipphawong J. Results of a double-blind, placebo-controlled, fixed-dose assessment of once-daily OROS hydromorphone ER in patients with moderate to severe pain associated with chronic osteoarthritis. Pain Pract. 2013;13(1):18-29. PMID: 22537100.

Rauck RL, Potts J, Xiang Q, et al. Efficacy and tolerability of buccal buprenorphine in opioid-naive patients with moderate to severe chronic low back pain. Postgraduate Medicine. 2016;128(1):1-11. PMID: 26634956.

Schwartz S, Etropolski M, Shapiro DY, et al. Safety and efficacy of tapentadol ER in patients with painful diabetic peripheral neuropathy: results of a randomized-withdrawal, placebo-controlled trial. Curr Med Res Opin. 2011;27(1):151-162. PMID: 21162697.

Serrie A, Lange B, Steup A. Tapentadol prolonged-release for moderate-to-severe chronic osteoarthritis knee pain: a double-blind, randomized, placebo- and oxycodone controlled release-controlled study. Curr Med Res Opin. 2017;33(8):1423-1432. PMID: 28537501.

Wen W, Sitar S, Lynch SY, et al. A multicenter, randomized, double-blind, placebo-controlled trial to assess the efficacy and safety of single-entity, once-daily hydrocodone tablets in patients with uncontrolled moderate to severe chronic low back pain. Expert Opin Pharmacother. 2015;16(11):1593-1606. PMID: 26111544.