

Extracting Key Messages from Systematic Reviews

TIMOTHY S. CAREY, MD, MPH
CATHY L. MELVIN, PHD, MPH
LEAH M. RANNEY, PhD

Systematic reviews have become a common method of synthesizing literature to examine the comparative effectiveness of medical interventions. Groups such as the Cochrane Collaboration and the U.S. Agency for Healthcare Research and Quality (AHRQ) conduct dozens of such reviews each year. However, these documents are often hundreds of pages long and the intended audiences, including medical providers and payers, may have difficulty interpreting the often technical terminology used in these reports. In this article, we describe the derivation of “key concepts” on the efficacy of antiepileptic drugs (AEDs) in the treatment of bipolar disorder from a previously published systematic review of the use of AEDs (AED Review) for multiple indications, including mood disorders and chronic pain. With the aid of a multidisciplinary science panel, we derived the key concepts from the source report and subsequent updates of that report. Because we found that the key concepts were still quite technical, we subsequently derived four less technical “key messages” and revised these through multiple additional iterations. The concepts and messages were then tested with key informants and focus groups. At all stages of the process, we found that it was critical to maintain fidelity to the initial systematic review. The structured approach used in the derivation process described here proved to be very helpful in developing key messages and concepts. (*Journal of Psychiatric Practice* 2007;14(suppl 1):28–34)

KEY WORDS: bipolar disorder, antiepileptic drugs, carbamazepine, gabapentin, lamotrigine, valproate, off-label use, systematic reviews

The science of evidence-based practice and systematic review has advanced substantially over the past decade. Systematic reviews are now done much more frequently and are updated more regularly; they are also more generally accepted. The evidence on which to base systematic reviews of clinical efficacy and effectiveness is much broader than in previous decades. Unfortunately, the translation of the science of systematic reviews into clinical practice has lagged behind.

Systematic reviews, conducted by experts in information science, clinical practice, and clinical epidemiology, are planned and written to rigorous quality standards. This results in reviews that are often very long and sometimes employ a substantial amount of jargon. The average practitioner is likely to have difficulty with long web-based or even academic journal-length reviews. Therefore, the translation of systematic reviews into evidence-based messages is a key intermediate step in producing changes in clinical practice.

Obviously, the process of developing and disseminating such evidence-based messages is not sufficient in itself to produce a change in practice, but this does represent a necessary step in the process of practice

improvement. Additional steps include use of appropriate reimbursement systems, administrative simplicity, technical and other support for practice change, and the availability of alternative treatment strategies.^{1–5}

This article describes the process by which key messages were derived from a systematic review. This process was part of a 3-year project, the goal of which was the national dissemination of fair and balanced information concerning use of antiepileptic drugs (AEDs) in the treatment of bipolar disorder. While variations of key messages derived from evidence reviews

CAREY, MELVIN, and RANNEY: The Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill.

Copyright ©2008 Lippincott Williams & Wilkins Inc.

Please send correspondence and reprint requests to: Cathy L. Melvin, PhD, MPH, The Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill, 725 Martin Luther King Jr. Boulevard, CB#7590, Chapel Hill, NC 27599-7590. cathy_melvin@unc.edu

Supported by a grant administered by the Vermont Office of the Attorney General and funded by a consortium of state attorneys general.

will likely have some unique characteristics specific to each project, we believe our experience is illustrative of the challenges inherent in the process itself. If the information contained in systematic reviews is to become useful and result in practice improvement, such a translation process is essential.

METHODS

The project described here was funded by the Neurontin Executive Committee, a consortium of state attorneys general (see Melvin, et al. in this supplement for a more detailed description of this project, p. 9⁶). The Neurontin Executive Committee contracted with the Cecil G. Sheps Center for Health Services Research at the University of North Carolina at Chapel Hill to develop a national dissemination program concerning the utilization of AEDs for bipolar disorder based on a systematic drug class review conducted as part of the Drug Effectiveness Review Project (DERP), which involved a series of drug class reviews coordinated by the Oregon Health Sciences University.^{7,8} The DERP is an established systematic review project conducting drug class reviews for a broad consortium of state Medicaid agencies. The original systematic review concerning the AEDs conducted by DERP was posted in December 2005 and the final report was completed in May 2006. The scope of the AED Review was quite broad and addressed use of AEDs for a variety of indications, including bipolar mood disorder, neuropathic pain, and fibromyalgia. The original AED Review, including evidence tables, was over 700 printed pages. Thus, the need to distill this voluminous evidence into messages that would be accessible to clinicians and policymakers was evident.

We first assembled a multi-disciplinary “science panel” that included experts in systematic review, psychiatry, and pharmacology. We were fortunately able to include the primary author of the original systematic review on the panel. All panelists disclosed any potential conflicts of interest and no relevant competing interests with pharmaceutical companies were identified. The panel was assisted by staff members. The initial task of the science panel was to read the report in detail and discuss the findings with the primary author. The panel also reviewed multiple background papers as well as pivotal trials of AEDs in bipolar disorder. Because the focus of the dissemination effort was the role of gabapentin in the treatment of bipolar disorder, the panel focused on primary publications concerning use of this agent in bipolar disorder.

Meetings, conference calls, and web-based sharing of documents greatly facilitated this process. Although the focus of the dissemination process was largely utilization of gabapentin, the panel members quickly recognized that simple cautionary messages regarding use of this medication would likely be perceived as essentially negative without giving physicians a more positive message as to therapeutic options and efficacious agents for the treatment of this serious and often life-long illness. The panel members were also reluctant to proceed immediately from the 700 page report to actionable messages; instead, they adopted an intermediate strategy of moving from the report to “key concepts.”

Another activity undertaken early in the project was the pharmaco-epidemiologic analysis of national databases in order to gather information concerning prescribing practices for AEDs in psychiatric disease. Based on this analysis, the staff easily concluded that the audience for the messages being developed would be practicing psychiatrists, since they prescribe the vast majority of AEDs for bipolar disorder.

The panel also found it necessary to expand somewhat on the data presented in the evidence report, since they wished to examine the strength of the evidence regarding the efficacy (or lack of efficacy) of the pharmacologic agents under consideration in more detail. In consultation with the staff of the UNC evidence-based practice center, the science panel used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system to evaluate this evidence.⁹ The GRADE system is a method for assessing the strength of evidence. It provides a global qualitative assessment, which is often operationalized as an assessment of the likelihood that additional work could lead to a different conclusion. Thus, the GRADE system includes high ratings (“further research is very unlikely to change the estimate of effect”), moderate ratings (“further research is likely to have important impact on the confidence in the estimate of effect and may change the estimate”), and low ratings (“further research is very likely to have an important impact on the estimate effect and is likely to change the estimate”). Some variations of the GRADE system also include a “very low” rating. GRADE ratings were done in parallel with generation of the initial draft key concepts from the evidence report. Over the course of 7 months, key concepts were drafted by staff, presented to the science panel, revised, redrafted, and presented again during multiple conference calls and face-to-face meetings. In total, eight successive versions of the key concepts were elaborated. These versions intentionally began at a very general

level and became successively more specific. However, at some point, the panel found that the specificity of many of the key concepts was too great and not sufficiently supported by evidence in the report. When this occurred, these key concepts were revised to be somewhat more general. The members of the science panel emphasized the importance of maintaining fidelity to the evidence throughout the process as they examined the strength of the findings concerning possible benefits that can be ascribed to the use of AEDs in the treatment of bipolar disorder. A final critical issue for the science panel was making a clear distinction between presentation of evidence in the form of key concepts and messages and development of a practice guideline. Essentially all the members of the panel had previously participated in guideline development activities and recognized that evidence is the raw material upon which guidelines should be based. Thus, the members of the panel felt strongly that the key messages being developed should describe the evidence and should not be prescriptive statements concerning what providers “should” or “should not” do in practice.

After multiple iterations of the key concepts had been developed, the panel then developed more simply worded messages derived from the somewhat more technical key concepts. These messages went through six major revisions in a process that once again required that the messages be frequently checked against the core literature of the original evidence report and sometimes against constituent trials and observational studies. Ensuring that the key concepts and messages maintained fidelity with the systematic review of best evidence was the top priority throughout the project.

Evidence reports and systematic reviews have a finite life span, since scientific evidence is in constant evolution. Thus, in the late fall of 2006, the AED Report was updated with additional trials,^{10,11} necessitating further review by the staff and the members of the science panel. These additional findings were then compared against the findings presented in the key concepts and messages. While no major changes were made in the key concepts and messages as a result of this update, several findings were reinforced.

Finally, the messages were tested using interviews with key informants and focus groups of practicing psychiatrists (for a description of this process, see Kish-Doto et al., p. 35¹²). The valuable feedback obtained from this process did not change the content but did alter the presentation and wording of the messages. Wording in the draft messages that was perceived as being more indicative of guideline development rather than mes-

sage dissemination was changed; the final messages were also more focused.

RESULTS

As discussed above, members of the science panel came from different disciplines, and the panel’s initial work reflected an excellent collegial relationship among the panel members characterized by respect for the specialized expertise of each of the other disciplines represented. Some of the work also required that members of the panel acquire specialized knowledge (for example, not all of the members were familiar with the GRADE rating system for strength of evidence at the outset of the project).

The final key concepts are listed in Table 1 and can also be accessed on our website (www.prescribingforbetteroutcomes.org). While the key concepts that were developed were of course much briefer than the complete evidence report, they were still relatively wordy and the reviewers thought the language was somewhat equivocal. For example, the wording of the first key concept in Table 1, which maintains fidelity to the underlying report, is quite dense and subject to more than one interpretation. As noted above, multiple iterations were required to arrive at this level of specificity. However, with regard to our primary purpose of describing the role of gabapentin in the treatment of bipolar disorder, the key concept related to that issue (see Concept 8 in Table 1) was more succinct.

Panel members thought that these key concept formulations were necessary in order to be clear regarding the acceptability of the evidence. For example, multiple case series have tried to describe the efficacy of gabapentin in bipolar disorder, but the science panel unanimously believed that such case series evidence was not acceptable in demonstrating efficacy of a treatment in a condition which has a waxing and waning course, such as bipolar disorder.

Although the equivocal nature of the key concepts was judged to be inappropriate for a broad dissemination campaign, the key concepts remain valuable as backup documentation and references for the multiple iterations of the key messages that were developed based on these concepts.

The science panel had the most difficulty describing the adverse events reported in association with the AEDs. Such issues are unfortunately ubiquitous in drug class reviews. Randomized trials of efficacy are often too small to allow calculation of rates of adverse events (also called “harms”) due to the medication. Pharmaco-

Table 1. Key Concepts about antiepileptic drugs (AEDs) derived from the Drug Effectiveness Review Project (DERP) Report^{*,7}

1. Current evidence supports the conclusion that three AEDs (carbamazepine, valproic acid/valproate, and lamotrigine) are efficacious in maintaining remission for adult outpatients with primary diagnoses of *bipolar I disorder with recent mania or mixed episodes*.
 - a. The overall magnitude of benefit obtained with AEDs in bipolar I disorder with recent mania or mixed episodes was an absolute improvement of the probability of attaining remission ranging from 7% to 28%; the relative rate of attaining remission was between 1.17 and 2.87 compared with placebo. The strength of evidence for this indication is low (Grading of Recommendations, Assessment, Development and Evaluation [GRADE] criteria).
 - b. Carbamazepine is the only AED that has been shown in fair-quality published trials to be significantly better than placebo in reducing mania scores in acute therapy of adult outpatients. Evidence is stronger for lamotrigine in prevention of depressive than manic episodes.
 - c. There was no acceptable evidence to support choice of one agent over another based on speed of onset in attaining remission.
2. Current evidence provides only modest support for the efficacy of the same three AEDs in achieving and maintaining remission in adult outpatients with *bipolar I disorder with a recent depressive episode or bipolar II disorder*. The strength of evidence in either of these subtypes of bipolar disorder is less than the strength of evidence regarding efficacy in bipolar I disorder with recent mania or mixed episodes.
 - a. The overall magnitude of benefit obtained with AEDs relative to placebo in bipolar I disorder with a recent depressive episode was an absolute improvement in attaining remission of 11%, with a relative rate of attaining remission of 1.44 compared with placebo. The GRADE strength of evidence for use of AEDs for this indication is low.
 - b. The overall magnitude of benefit in *bipolar II disorder* is an absolute difference of 15%, with a relative rate of attaining remission of 1.58 compared with placebo. The GRADE strength of evidence for use of AEDs for this indication is low.
3. Efficacy of these agents in maintaining remission is generally based on the percentage of patients who do not experience symptomatic recurrence or prematurely discontinue study treatment because of symptoms. There is great variability among the methods used to measure symptoms and functional status in these populations. The evidence indicating improvement in manic or depressive symptoms is much less clear due to difficulties comparing different symptom scores and functional status measures in these populations. Most available evidence does, however, demonstrate improvement in symptom scores compared with placebo. The level of absolute improvement in reducing recurrence ranged from 1% to 23%, with a risk reduction ranging from 0.63 to 0.96. The GRADE strength of evidence for use of AEDs for this indication is low.
4. Carbamazepine, valproic acid, and lamotrigine appear to have similar magnitudes of benefit in inducing remission, although the risk of recurrence is substantial for all agents. While the three drugs have similar magnitudes of benefit based on indirect comparisons, few studies directly compare these medications with each other; therefore, these conclusions should be considered tentative. The GRADE strength of evidence supporting this conclusion is low.
5. The rates of achieving and maintaining remission during treatment with the three AEDs mentioned above are similar to those obtained with lithium treatment for *bipolar I disorder*. For adult outpatients with acute mania, carbamazepine and valproate were similar, relative to lithium, in terms of response rates.
 - a. The incidence of recurrence in the studies examined ranged from 16% to 70% with placebo and from 6% to 65% with medication treatment. The broad range of these estimates is due to the variable definitions of recurrence used and variable durations of follow-up. Recurrence is a significant problem for these patients even with treatment with AEDs.

Table 1. continued

6. The overall risk of adverse events resulting in medication discontinuation is similar among the three AEDs listed above, and the overall risk of adverse events for AEDs is similar to that for lithium across all clinical subtypes of bipolar disorder. However, the types of adverse events encountered differ among the three AEDs and lithium. Serious adverse events, although uncommon, may occur with each agent; meaningful comparisons of the rates of these serious events among agents could not be performed. Overall evidence regarding comparative adverse event rates is based on a small number of studies.
 - a. Between 5% and 24% of patients discontinue an AED due to adverse events. The rates of serious adverse events are on the order of 0% to 10.3%. The GRADE strength of evidence for this finding is low.
 - b. Rates of skin reactions are higher for carbamazepine and lamotrigine; the most common predictor of rash is history of previous rash with an AED.
7. The risk of suicide or suicide attempt is present in bipolar disorder.
 - a. Compared with placebo, patients treated with AEDs for psychiatric conditions may have a small increase in suicidal ideation (<http://www.fda.gov/cder/drug/InfoSheets/HCP/antiepilepticsHCP.htm>).
 - b. Evidence regarding protection against the risk of suicide attempt is limited and does not support a difference between valproate and carbamazepine. Both of these agents have been associated with a lower protective effect against suicide attempt than lithium in observational studies; however, the evidence of less protective effect is stronger for valproate. There are insufficient data to compare the risk of suicide attempt among other AEDs.
8. There is some evidence (although not extensive) showing that gabapentin is no more, and perhaps less, efficacious than placebo in the treatment of bipolar I disorder with recent mania and rapid cycling bipolar disorder. No acceptable evidence was found to support the use of gabapentin in achieving remission or preventing relapse in bipolar disorder. Evidence regarding efficacy of topiramate for any of the conditions discussed above (bipolar I disorder with recent mania, hypomania, or mixed episodes and bipolar II disorder) is sparse. Current evidence regarding use of topiramate for acute mania shows no evidence of efficacy.
9. The available evidence regarding potential differential efficacy among AEDs in the treatment of patients with rapid cycling and other patient subgroups is extremely limited. Available evidence does not allow prediction of which patient subpopulations will respond to any given AED. Available evidence does not allow prediction of response to a subsequent AED based on response to an initial trial of therapy.
10. Little evidence is available regarding differential efficacy of the three AEDs in subpopulations defined by gender, age, ethnicity, or comorbidity.

**These key concepts are derived from the Drug Class Review on antiepileptic drugs in bipolar mood disorder. The population of interest is adults with new onset or chronic bipolar I and II disorder. While most of the studies in the review were conducted in outpatients, some of the studies involved patients with acute disease who were hospitalized for initiation of therapy.*

epidemiologic studies may be limited by lack of specificity regarding the nature of the adverse event and difficulty assigning a rating of causality. Data on adverse events occurring in association with medication are often limited to case reports, which make comparison of rates of adverse events among drugs very difficult, since there is no denominator available with which to calculate a rate. Pharmaco-epidemiology studies may allow calculation of a rate of adverse events from a drug, but

are often limited by lack of detail regarding patient characteristics and comorbidities. A particular issue with bipolar disorder is the relatively high rate of suicide attempts and completion. The panel discussed, at some length, this devastating consequence of the disease as well as the important distinctions between suicide as a consequence of the underlying disease and suicide as a consequence of medication. Rather than a consequence of medication, the panel felt that rates of

EXTRACTING KEY MESSAGES FROM SYSTEMATIC REVIEWS

Table 2. Key Messages about antiepileptic drugs (AEDs) derived from AED Key Concepts (Table 1)*

<i>Key Messages</i>	<i>Refers to Key Concept #</i>
1. There remains no scientifically acceptable clinical trial evidence which supports use of either gabapentin or topiramate in bipolar mood disorder, either as monotherapy or as an adjunct to other therapies.	8
2. Research supports the use of three AEDs—1) carbamazepine, 2) valproic acid/valproate, and 3) lamotrigine—in achieving and maintaining remission for adult outpatients with primary diagnoses of bipolar I disorder. Evidence of efficacy is less clear for these treatments for bipolar II disorder.	1, 2, 3, 4
3. Carbamazepine, valproic acid/valproate, and lamotrigine work as well as lithium in achieving and maintaining remission in bipolar I disorder, but the strength of the evidence supporting this conclusion is low, and additional research is needed to clarify the relative roles of these agents in bipolar disorder.	5
4. The types of adverse events vary among AEDs and lithium. There is insufficient evidence to determine if the overall risk of adverse events differs among AEDs. Unlike the AEDs, lithium poses a significant risk when taken in an overdose.	6, 7

**Key Concepts 9 and 10 refer to specific subpopulations that are not included in the Key Messages.*

suicidality were more likely a result of inadequate doses of medication, insufficient duration of treatment, or individual pharmacokinetics. However, the lack of large randomized trials made any conclusions regarding these issues very tentative.

Four key messages were eventually derived from the 10 key concepts (Table 2). Pre-testing with members of the practice community was especially valuable in refining the first and fourth key messages (see Kish-Doto et al. p. 35¹² and Melvin et. al. p. 44¹³ for a full description of the audience research). While practitioners often stated that they “didn’t use gabapentin” as a primary drug in the treatment of bipolar disorder, they often followed this up by saying that they sometimes used it as adjunctive rather than primary treatment. Therefore, the science panel added an explicit statement reflecting the lack of evidence for the efficacy of gabapentin either in primary or adjunctive use.

The fourth key message regarding adverse events was also modified. While the research literature showed no significant differences among the AEDs and lithium with regard to withdrawal from therapy or hospitalization for adverse events, clinicians consistently focused on the greater clinical impact of overdoses of lithium compared with overdoses of other agents such as the AEDs. The panel acknowledged this clinical insight and revised the fourth key message to address this issue.

DISCUSSION

Our experience in developing key messages based on a systematic review of the use of AEDs in bipolar disorder illustrates several of the challenges in disseminating the best evidence available for treatment of serious chronic conditions. Systematic reviews are certainly necessary for practice change since they summarize available evidence. As such, they serve as the necessary but not sufficient “science” base for the dissemination of best available treatments in practice. Moreover, even the relatively brief “key concepts” we derived from the systematic review were not the same as the key messages that we ultimately developed for dissemination. The process described in this article required the active participation of individuals with expertise in the methods and content of both psychiatry and systematic reviews as well as input from our intended audience of practicing psychiatrists and from experts in health communication and dissemination.

Modern systematic reviews, such as those commissioned by the Agency for Healthcare Research and Quality (www.ahrq.gov), address many key questions and often examine large classes of medications. We found that the science panel required at least several months to digest, interpret, and boil down one such extensive review into key concepts and messages.

The utilization of a multidisciplinary “science panel” throughout the process provided insights into the importance of obtaining the perspectives of multiple disciplines, including the end-user community. Similarly, early pre-tests of the messages were extremely useful and led to significant revisions. Even after revision, not all potential users agreed with the messages, but the panel recognized that transmission of evidence-based messages to individuals who might not necessarily agree with all of them is part of the process of practice quality improvement.

Maintaining fidelity to the original evidence report while deriving messages couched in practical straightforward terms represents a challenge to all translational researchers. Recently, the National Institutes of Health (NIH) launched a national Clinical Translational Science Award (CTSA) program. Translational research is of two types: “Type I” speeds the testing of novel therapies in humans—i.e., it involves translation from bench to bedside. “Type II” research, in contrast, seeks to speed the movement of efficacious therapies from clinical trial settings into practice. This second type of translational research requires assembling data through systematic reviews, since few clinical policy changes are based on a single clinical trial. The results of these trials and systematic reviews must be disseminated to clinicians and the public in clear terms that also discuss the limitations of the evidence base. Communication between the systematic review community and the practice community should occur early in the process. This second type of translation should not be viewed as information that is being “handed over” from systematic reviewers. Rather, the systematic review team must remain involved with the dissemination of the information by conducting appropriate updates and especially by maintaining fidelity between the messages and the underlying evidence. Although this fidelity can never be 100% precise at the first pass, ongoing communication among all parties will assure the best possible fidelity. Simply instructing clinicians to go to a website and absorb very bulky and technical reports will likely only result in the perpetuation of

practices that are not evidence-based or potentially useful.

References

1. Oxman AD, Thomson MA, David DA, et al. No magic bullets: A systematic review of 102 trials of interventions to improve professional practice. *CMAJ* 1995;153:1423–31.
2. Pathman DE, Konrad TR, Freed GL, et al. The awareness to adherence model of the steps to clinical guideline compliance. *Med Care* 1996;34:873–89.
3. Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: An overview of systematic reviews of interventions. *Med Care* 2001;39:112–45.
4. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:1–72, iii–iv.
5. Grimshaw JM, Eccles M, Thomas R, et al. Toward evidence-based quality improvement. Evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966–1998. *J Gen Intern Med* 2006;21:S14–20.
6. Melvin CL, Carey T, Oldham JM. Introduction: Antiepileptic medications for bipolar disorder. *J Psychiatr Pract* 2008;14 (suppl 1): 7–8.
7. Goodman F, Glassman P, Maglione M, et al. Drug class review on antiepileptic drugs in bipolar mood disorder, neuropathic pain, and fibromyalgia. Portland: Oregon Evidence-based Practice Center; 2006 (available online at either www.prescribingforbetteroutcomes.org or http://www.ohsu.edu/drugeffectiveness/reports/documents/_AED%20Final%20Report%20Update%201_unshaded.pdf).
8. Helfand M. Using evidence reports: Progress and challenges in evidence-based decision-making. *Health Aff* 2005;24:123–7.
9. Guyatt G, Vist G, Falck-Ytter Y, et al. An emerging consensus on grading recommendations? *ACP J Club* 2006;144:A8–9.
10. Kushner SF, Kahn A, Lane R, et al. Topiramate monotherapy in the management of acute mania: Results of four double-blind, placebo controlled trials. *Bipolar Disord* 2006;8:15–27.
11. Vieta E, Manuel-Goikolea J, Martinez-Aran A, et al. A double-blind, randomized controlled prophylaxis study of adjunctive gabapentin for bipolar disorder. *J Clin Psychiatry* 2006;67:473–7.
12. Kish-Doto J, Evans WD, Squire C, et al. Patterns of prescribing antiepileptic drugs for bipolar mood disorder. *J Psychiatr Pract* 2008;14 (suppl 1):35–43.
13. Melvin CL, Ranney L, Carey T, et al. Disseminating findings from a drug class review: Using best practices to inform prescription of antiepileptic drugs for bipolar disorder. *J Psychiatr Pract* 2008;14(suppl 1):44–56.