



# INSTITUTE FOR CIVIL JUSTICE and RAND HEALTH

THE ARTS  
CHILD POLICY  
CIVIL JUSTICE  
EDUCATION  
ENERGY AND ENVIRONMENT  
HEALTH AND HEALTH CARE  
INTERNATIONAL AFFAIRS  
NATIONAL SECURITY  
POPULATION AND AGING  
PUBLIC SAFETY  
SCIENCE AND TECHNOLOGY  
SUBSTANCE ABUSE  
TERRORISM AND  
HOMELAND SECURITY  
TRANSPORTATION AND  
INFRASTRUCTURE  
WORKFORCE AND WORKPLACE

This PDF document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

## Support RAND

[Purchase this document](#)

[Browse Books & Publications](#)

[Make a charitable contribution](#)

## For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore [RAND Institute for Civil Justice](#)  
and [RAND Health](#)

View [document details](#)

## Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents.

This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

# Evaluating Medical Treatment Guideline Sets for Injured Workers in California

---

Teryl K. Nuckols, Barbara O. Wynn, Yee-Wei Lim, Rebecca N. Shaw,  
Soeren Mattke, Thomas Wickizer, Philip Harber, Peggy Wallace,  
Steven M. Asch, Catherine MacLean, Rena Hasenfeld Garland

Prepared for the Commission on Health and Safety and Workers' Compensation and  
the Division of Workers' Compensation, California Department of Industrial Relations



INSTITUTE FOR CIVIL JUSTICE  
and RAND HEALTH

The research described in this report was conducted by the RAND Institute for Civil Justice and RAND Health, units of the RAND Corporation. This research was sponsored by the Commission on Health and Safety and Workers' Compensation and the Division of Workers' Compensation, California Department of Industrial Relations.

**Library of Congress Cataloging-in-Publication Data**

Evaluating medical treatment guideline sets for injured workers in California /  
Teryl K. Nuckols ... [et al.].

p. cm.  
"MG-400."

Includes bibliographical references.

ISBN 0-8330-3835-4 (pbk. : alk. paper)

1. Workers' compensation—Law and legislation—California. 2. Disability  
evaluation—Law and legislation—California. 3. Insurance, Disability—Medical  
examinations—California. I. Nuckols, Teryl K. II. Rand Corporation.

KFC592.E93 2005

344.79402'18—dc22

2005020346

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**RAND®** is a registered trademark.

© Copyright 2005 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2005 by the RAND Corporation

1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

1200 South Hayes Street, Arlington, VA 22202-5050

201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516

RAND URL: <http://www.rand.org/>

To order RAND documents or to obtain additional information, contact

Distribution Services: Telephone: (310) 451-7002;

Fax: (310) 451-6915; Email: [order@rand.org](mailto:order@rand.org)

## Summary

---

### Introduction

In recent years, the California workers' compensation system has been encumbered by rising costs and high utilization of medical care. Medical costs for injured workers grew by 111 percent between 1997 and 2002 and now represent more than half the total costs of workers' compensation (California Workers' Compensation Institute, 2004). Medical care payments were more than twice the national average in 2002 (National Academy of Social Insurance, 2004).

A comparative study across 12 states performed by the Workers' Compensation Research Institute concluded that California's higher medical costs resulted primarily from high utilization rather than high prices (Telles, Wang, and Tanabe, 2004). The study found that

- California had more visits per claim—in total and for physicians, chiropractors, and physical/occupational therapists—than any of the other states studied.
- The average number of visits for more-mature claims was 31 percent higher for hospitals, 70 percent higher for physicians, and 150 percent higher for chiropractors than the 12-state median.

To address these concerns, the California legislature passed a series of initiatives aimed at reducing costs and inappropriate medical care utilization in the system (AB 749 [Calderon], 2002; SB 228 [Alarcón], 2003; SB 899 [Poochigan], 2004). SB 228, passed in 2003, called for the adoption of medical treatment guidelines to define the appropriate utilization of medical care provided to injured workers, using the American College of Occupational and Environmental Medicine (ACOEM) guidelines as presumptively correct on an interim basis. Previously, physicians' treatment plans were presumed to be correct under the law. SB 899, passed in 2004, refined some of the requirements of SB 228. The study reported here, jointly sponsored by the California Commission on Health and Safety and Workers' Compensation (CHSWC) and the California Division of Workers Compensation (DWC), surveys and evalu-

ates medical treatment guidelines for injured workers in California, as specified in the revised labor code (California Labor Code, 2004):

§77.5(a): [CHSWC] shall conduct a survey and evaluation of evidence-based, peer-reviewed, nationally recognized standards of care.

§5307.27: [The Administrative Director of DWC, in consultation with CHSWC, will adopt after public hearings] a medical treatment utilization schedule, that shall incorporate the evidence-based, peer-reviewed, nationally recognized standards of care . . . and that shall address, at a minimum, the frequency, duration, intensity, and appropriateness of all treatment procedures and modalities commonly performed in workers' compensation cases.

In calling for guidelines specifying the appropriate utilization of medical care, SB 228 required CHSWC to survey and evaluate existing medical treatment guidelines. Using the results of the evaluation, the state was to adopt either the ACOEM guidelines or a better alternative in the longer term. By December 1, 2004, in consultation with CHSWC, the Administrative Director (AD) of DWC was required to adopt a utilization schedule based on CHSWC's recommendations (SB 228 [Alarcón], 2003).

## **Developing Research Objectives**

The legislation establishes a scientific basis for addressing medical care utilization in the California workers' compensation system. The phrase "evidence-based, peer-reviewed, nationally recognized standards of care" refers to the science of evidence-based medicine, which means using the best available research evidence to support medical professionals' decisionmaking (Sackett et al., 1996). The objective of evidence-based medicine has been defined as "minimizing the effects of bias in determining an optimal course of care" (Cohen, Stavri, and Hersh, 2004).

Medical treatment guidelines are an important tool for implementing evidence-based medicine. Guidelines are systematically developed statements that assist practitioner, patient, and, in this case, payor decisions about appropriate health care for specific clinical circumstances (Field and Lohr, 1990). A high-quality guideline can help curtail the effects of bias in formulating a treatment plan (AGREE Collaboration, 2001). Guidelines have many applications; perhaps the most common is distilling research evidence into a more usable form for busy clinicians. Insurers and third-party payors can also employ guidelines to determine whether a specific treatment is appropriate for a particular patient and therefore whether it should or should not be provided.

Techniques used by or on behalf of third-party payors to reduce health care costs by assessing the appropriateness of care provided to individual patients are collectively called *utilization management* (Gray and Field, 1989). There can be substantial variability in utilization management practices, particularly in the criteria used for assessing whether care is appropriate (Gray and Field, 1989; Wickizer and Lessler, 2002). Because a lack of standardization may affect access to and quality of care for patients, the recently passed workers' compensation legislation requires payors to employ review criteria that are consistent with the guidelines adopted by the state of California (California Labor Code, 2004).

To manage both the initial selection of treatment and the quantity of care provided, the adopted utilization schedule is required by SB 228 to address "frequency, duration, intensity, and appropriateness." Prior RAND researchers have defined appropriate medical care as care for which the potential benefits to the patient outweigh the potential risks, irrespective of cost. Inappropriate care is defined as care for which risks outweigh the potential benefits. Care of uncertain appropriateness falls between the two (Fitch et al., 2001). The current study used these existing definitions. The utilization schedule must also address, when relevant, frequency, intensity, and duration, i.e., quantity of care (SB 228 [Alarcón], 2003).

The legislation calls for guidelines addressing "all treatment procedures and modalities commonly performed in workers' compensation cases." Workers experience a broad range of injuries of the muscles, bones, and joints, as well as a wide variety of other medical problems. These often require diagnostic tests, such as X-rays and magnetic resonance imaging (MRI). In California, common therapies include medication, physical therapy, chiropractic manipulation, joint and soft-tissue injections, and surgical procedures.

To enable the state to manage medical utilization costs, the guidelines will have to address diagnostic tests and therapies that are not only common, but also costly, either individually or in the aggregate. Utilization management should be most cost-effective when it focuses on costly services (Wickizer, Lessler, and Franklin, 1999). Therefore, our analysis concentrated on diagnostic tests and therapies that are performed frequently and that contribute substantially to costs within the California workers' compensation system. We identified several such tests and therapies and consider them to be priority topic areas that the guidelines should cover: MRI of the spine, spinal injections, spinal surgeries, physical therapy, chiropractic manipulation, surgery for carpal tunnel and other nerve-compression syndromes, shoulder surgery, and knee surgery. Taken together, these procedures account for about 44 percent of the payments for professional services provided to California's injured workers. In addition, the surgeries account for about 40 percent of payments for inpatient hospital services.

## Guideline Evaluation Methods and Findings

Our study identified and evaluated guidelines for these priority areas. We first identified guidelines for work-related injuries; we then screened those guidelines, using multiple criteria; finally, we conducted comparative evaluations of the selected guidelines. It is important to note that we accomplished these objectives in a very limited time frame and with limited resources; because of these constraints, we did not conduct an independent review of the clinical literature, nor did we develop guidelines ourselves.

### Searching

We used the Institute of Medicine (IOM) definition of *guideline* as the basis for our search: “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (Field and Lohr, 1990). We also included documents developed to assist payor decisions, because the legislation called for the guidelines to address utilization issues.

Using a variety of complementary sources, we identified 72 relevant guidelines. We searched the National Library of Medicine’s MEDLINE and the National Guidelines Clearinghouse for practice guidelines published during the three years prior to June 2004, using keywords referring to work-related injuries. We surveyed the websites of relevant specialty society organizations listed by the American Medical Association (AMA). We contacted each of the other 49 U.S. states to inquire about workers’ compensation guidelines, and we interviewed national and California workers’ compensation experts, including providers, insurers, CHSWC and DWC staff, researchers, and our clinical panelists. We used Google to identify chiropractic guidelines and physical therapy guidelines, as well as to locate specialty society websites. We also posted a call for guidelines on the DWC website.

### Screening

We next began the task of selecting guidelines that satisfy the requirements of the legislation and preferences of the state (the criteria are listed in Table S.1). In accordance with the legislation, our first selection criterion was that the guidelines must be evidence-based and peer-reviewed. Our second criterion was that the guidelines must be nationally recognized. We developed generous definitions for these criteria in order to be inclusive at this stage. Together, *evidence-based* and *peer-reviewed* were taken to mean based, at a minimum, on a systematic review of literature published in medical journals included in MEDLINE. Systematic reviews of the literature are standard and essential features of an evidence-based guideline development process, as reflected by the fact that they are required by the National Guidelines Clearinghouse and are included in various guideline-assessment methodologies (AGREE

**Table S.1**  
**Screening Criteria for Guidelines Warranting Further Evaluation**

---

Evidence-based, peer-reviewed
Nationally recognized
Address common and costly tests and therapies for injuries of spine, arm, and leg
Reviewed or updated at least every three years
Developed by a multidisciplinary clinical team
Cost less than \$500 per individual user in California

---

Collaboration, 2001; National Guidelines Clearinghouse, 2004; Shaneyfelt, Mayo-Smith, and Rothwangl, 1999). *Nationally recognized* was taken to mean any one of the following: accepted by the National Guidelines Clearinghouse; published in a peer-reviewed U.S. medical journal; developed, endorsed, or disseminated by an organization based in two or more U.S. states; currently used by one or more U.S. state governments; or in wide use in two or more U.S. states.

The remaining criteria were developed in conjunction with CHSWC and DWC. Our third criterion was that guidelines must address, to at least a minimal degree, common and costly tests and therapies for injuries of the spine, arm, and leg. To address these tests and therapies, the state could (1) choose to have a universe of multiple acceptable guidelines addressing each topic; (2) choose the single best guideline for each topic, putting multiple guidelines together into a patchwork; or (3) choose one guideline set that addresses most or all of the topics. A universe of multiple guidelines would provide the most flexible decisionmaking for clinicians, whereas a patchwork would enable the state to choose the single highest-quality guideline for each topic and to expand the number of topics addressed.

We chose to evaluate sets of guidelines rather than multiple individual guidelines, for several reasons. Multiple guidelines may vary in rigor of development and frequency of updating. Moreover, they may address the same injuries and treatments and make contradictory recommendations, which could foster litigation. This is especially problematic for patients with multiple injuries, who might be subject to several different guidelines at the same time. Finally, multiple guidelines may be more complex for the state to implement and administer and may be costly to users. Of course, some of these problems could affect sets of guidelines as well, and the content within each set may vary in quality.

In hopes of identifying a single guideline set that would address many common and costly work-related injuries in a rigorous, evidence-based fashion and would also facilitate implementation, we decided to pursue the guideline-set approach at this point in time. The short timeline on this project precluded us from pursuing both this approach and the patchwork approach simultaneously. If no acceptable guideline sets could be identified, the state would have the option of considering alternative strategies in the future.

Our fourth selection criterion was that the guideline sets be reviewed at least every three years. This requirement was based on prior RAND research demonstrating that new research evidence renders about 50 percent of guidelines out of date after 5.8 years and at least 10 percent out of date after 3.6 years (Shekelle et al., 2001).

Our fifth criterion was that multidisciplinary clinical panels had to be involved in developing the guidelines. A 1990 IOM report on clinical practice guidelines considered a multidisciplinary development process to be an important component of guideline quality. The report asserted that use of a multidisciplinary team increases the likelihood that (1) all relevant scientific evidence will be considered, (2) practical problems with using the guidelines will be identified and addressed, and (3) affected [provider] groups will see the guidelines as credible and will cooperate in implementing them (Field and Lohr, 1990). Accepted guideline-assessment tools share the requirement for a multidisciplinary development process (AGREE Collaboration, 2001; Shaneyfelt, Mayo-Smith, and Rothwangl, 1999). Also, studies suggest that multidisciplinary panels produce more-balanced interpretations of the literature than single-specialty panels do (Coulter, Adams, and Skelelle, 1995). Finally, we believe that sets of guidelines addressing diverse therapies and injuries should have input from a variety of relevant experts.

Our sixth criterion was that guideline sets must cost less than \$500 per individual user. Some proprietary guidelines addressing work-related injuries are marketed predominantly to institutional users, such as insurers. In California, potential users of the workers' compensation medical treatment schedule also include providers, attorneys, judges, and many other types of individual users. We selected this threshold to ensure that evaluated guidelines would ultimately be available to individual as well as institutional users.

The following five guideline sets met all the screening criteria:

1. AAOS—Clinical Guidelines by the American Academy of Orthopedic Surgeons
2. ACOEM—American College of Occupational and Environmental Medicine Occupational Medicine Practice Guidelines
3. Intracorp—Optimal Treatment Guidelines, part of Intracorp Clinical Guidelines Tool<sup>®</sup>
4. McKesson—McKesson/InterQual Care Management Criteria and Clinical Evidence Summaries
5. ODG—Official Disability Guidelines: Treatment in Workers' Comp, by Work-Loss Data Institute

Many guidelines were eliminated because they did not address most of the cost-driver tests and therapies to at least a minimal degree. A few specialty society documents were excluded because they did not meet our definition of a guideline. Several

state guidelines and specialty society guidelines were eliminated because their content was out of date or because we could not confirm an updating plan. No guidelines were eliminated solely for lack of a multidisciplinary panel or on the basis of cost.

### Evaluating

The final step in our process was a comparative evaluation of the five selected guidelines, addressing both technical quality and clinical content. The technical quality evaluation assessed the process by which guidelines were developed and other dimensions. Although there are formal, accepted methods for developing guidelines, there is tremendous variation in the rigor of this process. We planned to exclude from further evaluation guidelines that performed especially poorly on technical quality. The clinical content evaluation assessed how well the guidelines address utilization decisions, i.e., appropriateness and quantity of treatment.

We evaluated technical quality with the AGREE instrument, which has been endorsed by the World Health Organization (WHO) and is becoming an accepted standard for guideline development (Grol, Cluzeau, and Burgers, 2003). AGREE addresses six domains that suggest an unbiased guideline (AGREE Collaboration, 2001):

1. **Scope and purpose:** whether the overall objective, clinical questions, and target patients are specifically described.
2. **Stakeholder involvement:** whether the developers had input from all the relevant professional groups, sought patients' preferences, and piloted the guideline among defined target users.
3. **Rigor of development:** whether developers used systematic and explicit methods to search for evidence and formulate recommendations, considered potential health benefits and risks, had the guideline externally reviewed, and provided an updating plan.
4. **Clarity and presentation:** whether the guideline makes specific and unambiguous recommendations, presents management options clearly, and includes application tools.
5. **Applicability:** whether developers considered organizational barriers and costs of applying the guideline and provided key review criteria for monitoring implementation.
6. **Editorial independence:** whether the guideline is editorially independent from the funding body and conflicts of interest of guideline development members have been recorded.

The RAND team rated the guideline sets on these domains, using the guidelines themselves as well as detailed descriptions and corroborating evidence provided by guideline developers.

All five of the selected guideline sets performed reasonably well in the technical evaluation, which produced standardized domain scores ranging from 0.00 (lowest) to 1.00 (highest) (Table S.2). *Scope and purpose* were well defined for all. *Stakeholder involvement* was weakest for AAOS, strongest for McKesson, and good for the rest. *Rigor of development* was very good for all. *Clarity and presentation* were excellent for all. *Applicability* was variable because developers often neglected implementation—McKesson was good, ODG was better, and the others were poor. *Editorial independence* was lowest for Intracorp and excellent for the rest.

Two prior studies that evaluated a total of about 150 guidelines found highly variable scores across all six domains (Burgers et al., 2004; Harpole et al., 2003). Our five selected guideline sets scored higher in the *rigor of development* and *editorial independence* domains than many guidelines did in other studies. Like guidelines from other studies, our five guidelines were relatively weak in the *stakeholder involvement* and *applicability* domains. Overall, the scores of our five guidelines were higher than those in the two prior studies, probably because we included additional details provided by guideline developers. Because all five of these guidelines did reasonably well in the technical quality evaluation, we decided none warranted elimination on this basis.

Next, a multidisciplinary clinical panel evaluated guideline content, assessing relevant content within each guideline and considering ten selected therapies in slightly greater detail. Relevant content addressed utilization decisions—specifically, appropriateness of care and quantity of care. We believe that, to be useful in making utilization decisions, the relevant content should be comprehensive (applicable to most patients) and valid (consistent with evidence or expert opinion). Panelists rated guidelines independently, then met to discuss areas of disagreement and to re-rate the guidelines.

For our panel, we selected 11 clinicians referred by national specialty societies. We sought national experts in musculoskeletal injuries who were practicing at least 20 percent of the time and who had some experience treating injured workers. Eight

**Table S.2**  
**Technical Quality Evaluation—AGREE Instrument Results**  
**(Standardized Domain Scores)**

Domain	AAOS	ACOEM	Intracorp	McKesson	ODG
Scope and purpose	1.00	0.89	0.89	1.00	1.00
Stakeholder involvement	0.54	0.79	0.79	0.88	0.79
Rigor of development	0.81	0.88	0.83	0.88	0.81
Clarity and presentation	0.96	0.88	1.00	1.00	0.96
Applicability	0.17	0.33	0.33	0.61	0.72
Editorial independence	1.00	1.00	0.75	1.00	0.92

national societies, representing a broad spectrum of providers caring for injured workers, made nominations. The only desired specialty that was not represented among our nominees was radiology. We selected clinical leaders from a diversity of geographic locations and practice settings, with diverse experience in caring for injured workers. To avoid potential conflicts of interest, we wanted no more than about 20 percent of the selected panelists to be from California, and we would have excluded panelists involved in the development of the guidelines under review. We gave preference to individuals experienced in the development, evaluation, or implementation of medical treatment guidelines, and experience with expert panels was a plus. To increase the discussion related to services not commonly ordered or provided by other panel members, we included two panelists expert in these services. We interviewed the most promising candidates by telephone to clarify their experience, and we contacted references to explore the ability of the candidates to function in groups. The final panel included one general internal medicine physician, two occupational medicine physicians, one physical medicine and rehabilitation physician, one physical therapist, one neurologist who is also board-certified in pain management, two doctors of chiropractic medicine, two orthopedic surgeons, and one neurosurgeon.

Panelists reviewed each guideline set in its entirety and evaluated ten selected therapies in detail: physical therapy, chiropractic manipulation, surgical decompression procedures, and surgical fusion procedures for lumbar spine problems; physical therapy, chiropractic manipulation, and surgery for carpal tunnel syndrome; physical therapy, chiropractic manipulation, and surgery for shoulder injuries. We selected therapies representing regions of the body frequently injured at work, such the spine and the large and medium-sized joints in the arms and legs. Within each category, we focused on cost-driver tests and therapies, preferring those for which the guidelines had different recommendations and for which we had panel nominees providing the services addressed. Our limited time frame forced us to narrow the number of topics under consideration. Because all of the guidelines made similar recommendations about spinal MRI and knee surgery, there seemed little benefit to comparing these topics. Furthermore, the lack of a radiologist on the panel would have made it difficult to evaluate MRI of the spine or spinal injections. This left us with the ten therapies listed above, which included surgery and physical modalities, i.e., physical therapy and chiropractic manipulation. We needed to distinguish between physical therapy and chiropractic manipulation because we did not want panelists to rate the same content twice. California chiropractors told us that there is some overlap between the physical modalities provided by these two specialties and that the appropriateness of manipulation influences chiropractors' decisions to provide other physical modalities. We therefore defined physical therapy as treatments provided by physical therapists and chiropractic manipulation as any additional treatments that can be provided only by chiropractors.

Although the residual (i.e., nonselected) content within each guideline varied in scope, we wanted to evaluate such content. Panelists rated residual content in each guideline as though it were a separate topic, considering other common and costly therapies for work-related injuries. Panelists also evaluated the entire content of each guideline, considering common and costly therapies for work-related injuries. They then rated and ranked the guidelines.

To facilitate rating, we provided the panelists with booklets containing relevant guideline chapters for the ten selected therapies, annotated to identify content addressing surgery, physical therapy, and chiropractic manipulation. For the residual- and entire-content evaluations, each panelist was provided with electronic access to the entire content of the five guidelines.

Because we identified no existing methods for rating the clinical content of guidelines, we adapted the RAND/UCLA Appropriateness Method (RAM), having panelists rate guideline comprehensiveness and validity for each of the various topics. Panelists rated comprehensiveness and validity separately on nine-point scales, with 9 as the highest rating. Panelists who were unfamiliar with a topic were instructed to rate the content a 5 (Fitch et al., 2001).

In the analysis, ratings were interpreted as follows:

- Comprehensive or valid: a median rating of 7 to 9 without disagreement.
- Not comprehensive or invalid: a median rating of 1 to 3 without disagreement.
- Uncertain comprehensiveness or validity: a median rating of 4 to 6, or any rating with disagreement.

After the panelists ranked the entire content of each guideline, we determined its median rank.

Using these methods, we found that the appropriateness of particular kinds of surgery is addressed well by the various guideline sets, as shown in Table S.3. In the table, *Yes* means the panel agreed that the content was both comprehensive and valid. *Not comprehensive* means the panel agreed that the guideline was not comprehensive; we assume minimal relevant content and do not report validity. *Not valid* means that the content was of uncertain or better comprehensiveness, and the panel agreed that the content was not valid. *Validity uncertain* means that the content was of uncertain or better comprehensiveness and the panelists were uncertain of validity.

Panelists agreed that the AAOS guideline set was valid and comprehensive for lumbar spinal decompression and fusion surgeries. They were uncertain whether it was valid for carpal tunnel surgery and agreed that it was not comprehensive in addressing shoulder surgery. Panelists agreed that the ACOEM guideline was valid and comprehensive for lumbar spinal decompression surgery, carpal tunnel surgery, and shoulder surgery. Validity was uncertain for lumbar spinal fusion surgery. Panelists

**Table S.3**  
**Panelists' Assessment of the Comprehensiveness and Validity of Content Addressing the Appropriateness of Surgical Procedures**

	AAOS	ACOEM	Intracorp	McKesson	ODG
Lumbar spinal decompression	<b>Yes</b>	<b>Yes</b>	Validity uncertain	<b>Yes</b>	Validity uncertain
Lumbar spinal fusion	<b>Yes</b>	Validity uncertain	Not valid	Validity uncertain	Validity uncertain
Carpal tunnel surgery	Validity uncertain	<b>Yes</b>	Validity uncertain	<b>Yes</b>	<b>Yes</b>
Shoulder surgery	Not comprehensive	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>

agreed that the Intracorp guideline was valid and comprehensive for shoulder surgery and invalid for lumbar spinal fusion surgery; the other two topics were of uncertain validity. The McKesson guidelines for surgical topics were rated the same as the ACOEM guidelines. The ODG guideline set was rated comprehensive and valid for both carpal tunnel surgery and shoulder surgery; the other two topics were of uncertain validity.

As shown in Table S.4, appropriateness of physical modalities is rarely addressed well by any of the five guidelines. Panelists were uncertain of the validity of the AAOS guideline for two topics and agreed that it was not comprehensive for the four others. Panelists agreed that the ACOEM guideline was valid and comprehensive for physical therapy of the shoulder. They agreed that it was not comprehensive for chiropractic manipulation of the shoulder. Validity was uncertain for the other four topics. Panelists agreed that the Intracorp guideline was not valid for chiropractic manipulation of the spine and carpal tunnel. Validity was uncertain for the remain-

**Table S.4**  
**Panelists' Assessment of the Comprehensiveness and Validity of Content Addressing the Appropriateness of Physical Modalities**

	AAOS	ACOEM	Intracorp	McKesson	ODG
Lumbar spine physical therapy	Validity uncertain				
Lumbar spine chiropractic	Not comprehensive	Validity uncertain	Not valid	Validity uncertain	Validity uncertain
Carpal tunnel physical therapy	Not comprehensive	Validity uncertain	Validity uncertain	Validity uncertain	<b>Yes</b>
Carpal tunnel chiropractic	Not comprehensive	Validity uncertain	Not valid	<b>Yes</b>	<b>Yes</b>
Shoulder physical therapy	Validity uncertain	<b>Yes</b>	Validity uncertain	<b>Yes</b>	Validity uncertain
Shoulder chiropractic	Not comprehensive	Not comprehensive	Validity uncertain	Not comprehensive	Not comprehensive

ing topics. They agreed that the McKesson guideline was valid and comprehensive for chiropractic manipulation of the carpal tunnel and physical therapy of the shoulder. They also agreed that it was not comprehensive in addressing chiropractic manipulation of the shoulder. Validity was uncertain for the other three topics. Panelists agreed that the ODG guideline was valid and comprehensive for physical therapy and chiropractic manipulation of the carpal tunnel. They agreed that it was not comprehensive in addressing chiropractic manipulation of the shoulder. Validity was uncertain for the other three topics.

Quantity of physical modalities is rarely addressed well by any of the five guidelines, as is evident from Table S.5. Panelists agreed that the AAOS guideline was not comprehensive in addressing the six quantity topics. They agreed that the ACOEM guideline was valid and comprehensive for physical therapy of the carpal tunnel. They agreed that it was valid for physical therapy of the shoulder but were uncertain of its comprehensiveness. Validity was uncertain for physical therapy of the spine. Panelists agreed that it was not comprehensive for the remaining three topics. Panelists agreed that the Intracorp guideline was not valid for chiropractic manipulation of the spine and carpal tunnel. It was of uncertain validity for all physical therapy topics and for chiropractic manipulation of the shoulder. Panelists agreed that the McKesson guideline was comprehensive and valid for chiropractic manipulation of the carpal tunnel. They agreed that it was not comprehensive for chiropractic manipulation of the shoulder. Validity was uncertain for the remaining topics. They agreed that the ODG guideline was comprehensive and valid for physical therapy of the shoulder, and they agreed that it was not comprehensive for chiropractic manipulation of the shoulder. Validity was uncertain for the remaining topics.

**Table S.5**  
**Panelists’ Assessment of the Comprehensiveness and Validity of Content Addressing the Quantity of Physical Modalities**

	AAOS	ACOEM	Intracorp	McKesson	ODG
Lumbar spine physical therapy	Not comprehensive	Validity uncertain	Validity uncertain	Validity uncertain	Validity uncertain
Lumbar spine chiropractic	Not comprehensive	Not comprehensive	Not valid	Validity uncertain	Validity uncertain
Carpal tunnel physical therapy	Not comprehensive	Not comprehensive	Validity uncertain	Validity uncertain	Validity uncertain
Carpal tunnel chiropractic	Not comprehensive	<b>Yes</b>	Not valid	<b>Yes</b>	Validity uncertain
Shoulder physical therapy	Not comprehensive	<b>Valid, comprehensiveness uncertain</b>	Validity uncertain	Validity uncertain	<b>Yes</b>
Shoulder chiropractic	Not comprehensive	Not comprehensive	Validity uncertain	Not comprehensive	Not comprehensive

Table S.6 presents summary results for each guideline, reiterating the appropriateness ratings, then presenting the residual-content and entire-content evaluations. To summarize, the panel ratings indicate that the panelists thought all five guideline sets require substantial improvement. However, they preferred the ACOEM guidelines.

1. The AAOS guideline addressed appropriateness well for two of the four surgical topics and none of the six physical modality topics. Panelists agreed that the guideline had little residual content. In the entire-content rating, panelists agreed the guideline was valid but were uncertain whether it was comprehensive. It was ranked last.
2. The ACOEM guideline addressed appropriateness well for three of the four surgical topics and one of the six physical modalities. Panelists were uncertain whether the residual content was valid. In the entire-content rating, panelists agreed that the guideline was valid but were uncertain whether it was comprehensive. It was ranked first.
3. The Intracorp guideline addressed appropriateness well for one of the four surgical topics and none of the six physical modalities. Panelists were uncertain whether the residual content was valid. In the entire-content rating, panelists agreed that the guideline was not valid. It was ranked third.
4. The McKesson guideline addressed appropriateness well for three of the four surgical topics and two of the six physical modalities. In the residual-content and

**Table S.6**  
**Clinical Evaluation Summary: Panelists' Assessment of Comprehensiveness and Validity**

	AAOS	ACOEM	Intracorp	McKesson	ODG
<b>Appropriateness</b>					
Surgery	2 of 4 topics	3 of 4 topics	1 of 4 topics	3 of 4 topics	2 of 4 topics
Physical therapy and chiropractic	0 of 6 topics	1 of 6 topics	0 of 6 topics	2 of 6 topics	2 of 6 topics
<b>Residual Content</b>					
	Not comprehensive	Validity uncertain	Validity uncertain	Validity uncertain	Validity uncertain
<b>Entire Content</b>					
Rating	<b>Valid</b> , comprehensiveness uncertain	<b>Valid</b> , comprehensiveness uncertain	Not valid	Validity uncertain	Validity uncertain
Median rank	4	<b>1</b>	3	2	2

entire-content evaluations, panelists were uncertain of validity. This guideline set tied for second.

5. The ODG guideline addressed appropriateness well for two of the four surgical topics and two of the six physical modalities. In the residual-content and entire-content evaluations, panelists were uncertain of validity. This guideline set tied for second.

Panelists' qualitative comments and discussion tone and content during the meeting were informative in interpreting these results. They appeared quite comfortable rating the surgical topics, based on their personal understanding of the relevant literature. However, for the physical modalities, panelists providing those services and those not providing them had quite different understandings. Some of the physicians were relatively unfamiliar with certain physical modalities, such as chiropractic manipulation of the carpal tunnel and shoulder. Providers of physical modality services cited published literature for their specialties, and physicians occasionally admitted being unfamiliar with that literature. For some physical modality topics, it appears that little literature may exist at this time. For example, the two chiropractors on the panel, both very familiar with evidence-based medicine and chiropractic guidelines, were aware of only two preliminary studies addressing chiropractic manipulation for carpal tunnel syndrome.

At the conclusion of the meeting, panelists elaborated upon their ratings and preferences. Several panelists voiced the opinion that all five guidelines require substantial improvement. Seven of the 11 panelists felt that

- The five selected guidelines “are not as valid as everyone would want in a perfect world.”
- “They do not meet or exceed standards; they barely meet standards.”
- “California could do a lot better by starting from scratch.”

Some panelists reported preferring the specialty society guidelines to the proprietary ones marketed for utilization management purposes, which they found too “proscriptive,” meaning that the proprietary guidelines limited clinical options to a degree that made the panelists uncomfortable.

The panelists' comments may shed light on some internal inconsistencies in our findings. One notable inconsistency is that the ACOEM and McKesson guidelines performed similarly for the selected topics and for the residual content, yet the ACOEM was judged valid overall and the McKesson was not. When asked about this, some panelists explained that the McKesson guideline was overly proscriptive, as noted above. Clinicians may be biased against guidelines marketed for utilization management purposes or biased in favor of specialty society guidelines. Alternatively,

the McKesson guideline may be overly proscriptive, limiting care options to an unacceptable degree.

Another inconsistency is the fact that all five guidelines did reasonably well in the technical quality evaluation, yet ratings were very uneven in the clinical content evaluation. This inconsistency was most pronounced for the physical modalities. There could be several possible explanations for this. First, even rigorously developed guidelines use expert opinion to fill gaps in the evidence. Such gaps appear common for physical modality issues, particularly quantity of care and chiropractic manipulation of the carpal tunnel. Panelists were less likely to agree that opinion-based recommendations are valid. Second, physicians might not know that chiropractors manipulate the extremities, making it difficult for them to develop or assess guidelines for such modalities. Third, although one would expect that good technical quality, including rigorous development methods, would produce valid clinical content, we know of no studies addressing this.

Our methods have important limitations that might also explain the inconsistencies. First, we were unable to provide panelists with literature reviews for the therapies under consideration. This is an especially important limitation for our evaluations of the physical modalities, because panelists understood this literature differently; and for chiropractic manipulation of the carpal tunnel, some panelists were not familiar with the relevant literature at all. Second, in typical RAND/UCLA appropriateness studies, panelists assess appropriateness for well-defined surgeries and categories of patients (Fitch et al., 2001). In contrast, we aggregated large amounts of clinical material and asked panelists to provide summary judgments. This may mean that panelists averaged highly valid content with invalid content, leading to intermediate, i.e., uncertain, summary judgments. The residual-content evaluation involved aggregating the largest amount of content; therefore, this weakness would be most pronounced in that evaluation. The residual content was rated of uncertain validity for four of the five guidelines. Third, to our knowledge, no methods for evaluating clinical content have been validated to date. We borrowed from validated methods to the degree possible, but the main premise of our evaluation, using an expert panel to assess and compare multiple guidelines, has not been described in the published literature.

Despite these limitations, the clinical content evaluation leads us to the following research conclusions. All five guideline sets appear far less than ideal—in the words of the panelists, they barely meet standards. The clinical panel preferred the ACOEM guideline to the alternatives and considered it valid but not comprehensive in the entire-content rating. The ACOEM guideline addresses cost-driver surgical topics and addresses them well for three of the four therapies the panel rated. A surgical weakness in the ACOEM guideline set, lumbar spinal fusion, is well addressed by the AAOS guideline set. The ACOEM guideline does not appear to address

physical modalities in a comprehensive and valid fashion, but the other four guidelines do little better. The same is true of the residual content in each guideline.

## Stakeholder Experiences and Insights

Since March 31, 2004, the ACOEM guideline has been implemented in the California workers' compensation system as presumptively correct on an interim basis. Through interviews with stakeholders, we learned about difficulties that have arisen during this period. Payors appear to be interpreting and applying the guideline inconsistently. Moreover, payors appear uncertain about which topics ACOEM covers in enough detail to determine appropriateness of care. Sometimes the guideline has been applied to topics that it addresses minimally or not at all, including chronic conditions, acupuncture, medical devices, home health care, durable medical equipment, and toxicology.

We received additional stakeholder input on the use of medical treatment guidelines within the California workers' compensation system after the clinical evaluation of the five guideline sets was completed. We invited selected stakeholders to a meeting, the purpose of which was twofold: to share our findings to date and to obtain their input on implementation issues. Most of the participants were representatives of stakeholder organizations that were suggested to us by CHSWC and represented a variety of perspectives: labor, applicants' attorneys, physicians and other practitioners, payors, and self-insured employers. Much of the meeting was spent on the issue of how the AD of DWC could address the topical areas in the ACOEM guidelines that need improvement.

A commonly shared viewpoint among the participants was that the longer-term goal should be to take the best guideline available for each topic area and patch these guidelines together into a single coherent set, but there were differing viewpoints on the mechanism for reaching that goal and the policies that should be adopted in the interim. Payors tended to favor "staying the course" until a more valid and comprehensive set could be developed. They noted that the ACOEM guidelines had just been implemented and that additional time was needed both to work out the issues with ACOEM and to consider carefully the consistency and administrative issues that might arise in using multiple guidelines. Other participants tended to favor using guidelines from different developers to address the shortcomings. They suggested different short-term strategies, ranging from using the AAOS guidelines for spinal surgery to adopting multiple guidelines for additional topical areas as long as they met some minimum criteria, such as listing in the National Guideline Clearinghouse or having been developed by the specialty societies. Longer-term strategies involved evaluating existing guidelines for other topical areas and working toward a comprehensive, consistent guideline set, using a multidisciplinary group of evaluat-

ors. These participants were concerned about the potential detrimental impact on workers of using guidelines with uncertain validity.

Because all of the comprehensive guideline sets we evaluated were of uneven quality, we agree with the common view among stakeholders that the state will need to patch multiple guidelines together into a coherent set. However, issues arise when multiple guidelines addressing the same topic are considered presumptively correct under the law. Identifying and resolving conflicting recommendations would therefore be helpful. Having a single high-quality guideline for each topic rather than multiple guidelines would probably minimize such conflicts.

On the basis of our research conclusions and the stakeholder comments described above, we make the following recommendations for the short term, the intermediate term, and the longer term.

#### **Short Term (After December 1, 2004)**

1. The panelists preferred the ACOEM guideline set to the alternatives, and this set is already in use in the California workers' compensation system; therefore, there is **no reason to switch to a different comprehensive guideline set at this time.**
2. ACOEM content was rated comprehensive and valid for three of the four surgical topics considered, and our evaluation methods appeared successful for these topics; therefore, **the state can confidently implement the ACOEM guidelines for carpal tunnel surgery, shoulder surgery, and lumbar spinal decompression surgery.**
3. Because spinal fusion surgery is especially controversial and risky, and its use is rapidly increasing in the United States (Deyo, Nachemson, and Mirza, 2004; Lipson, 2004), it warrants additional emphasis. The AAOS content was rated comprehensive and valid for this procedure and also for lumbar spinal decompression surgery. Therefore, **the state can confidently implement the AAOS guideline for lumbar spinal fusion surgery and, if convenient, for lumbar spinal decompression surgery.**
4. The ACOEM guideline set performed well for three of the four categories of surgery we evaluated. Generalizing these findings to other surgical topics would be reasonable; therefore, **the state could implement the ACOEM guideline for other surgical topics.**
5. We found the validity of the ACOEM guideline for the physical modalities and the remaining content uncertain, but our evaluation methods appeared to have important limitations for these areas; therefore, **we are not confident that the ACOEM guideline is valid for nonsurgical topics.** Deciding whether or not to continue using ACOEM for nonsurgical topics as an interim strategy remains a policy matter.

- a. We recommend that to identify high-quality guidelines for the nonsurgical topics, the state should proceed with the intermediate-term solutions described below as quickly as possible.
6. **We suggest implementing regulations to clarify the following:**
  - a. Stakeholder interviews suggest that payors in the California workers' compensation system are applying the ACOEM guidelines inconsistently, sometimes for topics the guidelines do not address or address only minimally; therefore, **we recommend that the state issue regulations clarifying the topics for which the adopted guidelines should apply.**
    1. Our stakeholder interviews suggest that acupuncture, chronic conditions, and other topics may not be covered well by the ACOEM guideline.
  - b. **For topics to which the adopted guideline does not apply, the state should clarify who bears the burden of proof for establishing appropriateness of care.**
  - c. **For topics that are not covered by the adopted guideline and throughout the claims adjudication process, the state should consider testing the use of a defined hierarchy to weigh relative strengths of evidence.**
  - d. Because the medical literature addressing appropriateness and quantity of care may be very limited for some physical modalities and other tests and therapies, some guideline content will include a component of expert opinion; therefore, **the state should clarify whether expert opinion constitutes an acceptable form of evidence** within "evidence-based, peer-reviewed, nationally recognized standards of care."
  - e. Our stakeholder interviews suggest that payors are uncertain whether they have the authority to approve exceptions to the guidelines for patients with unusual medical needs. Therefore, **the state should consider specifically authorizing payors to use medical judgment in deciding whether care at variance with the adopted guidelines should be allowed.**

### Intermediate Term

1. If the state wishes to develop a patchwork of guidelines addressing work-related injuries, our research suggests the following priority topic areas: **physical therapy of the spine and extremities, chiropractic manipulation of the spine and extremities, spinal and paraspinal injection procedures, MRI of the spine, chronic pain, occupational therapy, devices and new technologies, and acupuncture.**
  - a. When guidelines within a patchwork have overlapping content, the state may want to identify and resolve conflicting recommendations before adopting the additional guidelines.
2. Because high scores in the technical evaluation were not associated with high evaluations by expert clinicians, **we recommend that future evaluations of existing medical treatment guidelines include a clinical evaluation component.**

Specifically, we recommend against adopting guidelines solely on the basis of acceptance by the National Guideline Clearinghouse or a similar standard because this ensures only technical quality.

3. If the state wishes to employ the clinical evaluation method we developed for multiple future analyses, **we suggest that at least one analysis should involve an attempt to confirm the validity of the clinical evaluation method**, including determining the effect of a literature review on panel findings.
4. Lack of a comprehensive literature review appeared to be a major limitation in our evaluation of content addressing the physical modalities; therefore, **future evaluations addressing the physical modalities should include a comprehensive literature review.**

### Longer Term

1. Our technical evaluation revealed that ACOEM and AAOS developers did a poor job of considering implementation issues, and our stakeholder interviews indicated that payors are applying the ACOEM guideline in an inconsistent fashion. Therefore, **we recommend that the state develop a consistent set of utilization criteria (i.e., overuse criteria) to be used by all payors.**
  - a. Rather than covering all aspects of care for a clinical problem, as guidelines do, the utilization criteria should be targeted to clinical circumstances relevant to determining the appropriateness of specific tests and therapies.
  - b. Rather than defining appropriateness for all tests and therapies provided to injured workers, the criteria should focus on common injuries that frequently lead to costly and inappropriate services.
  - c. The utilization criteria should be usable for either prospective or retrospective assessments of appropriateness, because utilization management in the California workers' compensation system involves both types of activities.
  - d. The criteria should use precise language so that they will be interpreted consistently.
2. Another task within this project addresses developing a quality-monitoring system for California workers' compensation. Underuse of medical care is one important component of quality; therefore, the state may need to develop criteria for measuring underuse. **Developing the overuse and underuse criteria at the same time would be resource-efficient.**
3. There are two basic ways the state could develop overuse and underuse criteria:
  - a. **The criteria could be developed from existing guidelines**, such as the ACOEM, AAOS, and any other guidelines judged valid in future studies. We suspect that it may be somewhat difficult to develop overuse criteria from clinical guidelines.
  - b. **The criteria could be developed from the literature and expert opinion**, without the intermediate step of developing or selecting guidelines.