RAND LABOR AND POPULATION

Technical Quality and Clinical Acceptability of a Utilization Review Guideline for Occupational Conditions

ODG® Treatment Guidelines by the Work Loss Data Institute

Teryl K. Nuckols, Kanaka Shetty, Laura Raaen, Dmitry Khodyakov

For more information on this publication, visit www.rand.org/t/RR1819

Published by the RAND Corporation, Santa Monica, Calif.

Preface

Due to concerns about rising healthcare costs and the use of healthcare services that offer patients limited benefits, payers that cover occupational conditions are increasingly employing utilisation review techniques to assess the appropriateness of care provided to individual patients. Guidelines are frequently used as tools in utilisation reviews.

Since 2004, the number of guidelines used in workers' compensation settings that address multiple diverse occupational conditions has declined. The Official Disability Guidelines (ODG®) Medical Treatment Guidelines by the Work Loss Data Institute (WLDI) is a utilisation review guideline that is designed specifically for and widely used in the field of workers' compensation.

In recent years, stricter standards have been issued for guideline development to assure that guidelines are based on the best available evidence, are viewed by clinicians as valid, and that implementation will lead to improved patient outcomes. The objective of this work was to perform an independent evaluation of the ODG to assess technical quality (the rigour of development methods) as well as clinical acceptability (the validity of the guideline in the eyes of diverse clinical experts).

This research was undertaken within RAND Labor and Population and was supported by Insurance and Care New South Wales (icare NSW), a public financial enterprise governed by an independent board of directors that delivers insurance and care services to the people of New South Wales, Australia, including people with injuries under the NSW Workers Compensation Schemes.

RAND Labor and Population has built an international reputation for conducting objective, high-quality, empirical research to support and improve policies and organisations around the world. Its work focuses on children and families, demographic behaviour, education and training, labour markets, social welfare policy, immigration, international development, financial decisionmaking, and issues related to ageing and retirement with a common aim of understanding how policy and social and economic forces affect individual decisionmaking and human well-being.

For more information on RAND Labor and Population, contact
Unit Director
RAND Labor and Population
1776 Main Street, P.O. Box 2138
Santa Monica, CA 90407-2138
(310) 393-0411
or visit the Labor and Population homepage at:
http://www.rand.org/labor.

Summary

Due to concerns about rising healthcare costs and the use of healthcare services that offer limited benefit to patients, workers' compensation payers often employ utilisation review techniques. The Official Disability Guidelines (ODG®) Medical Treatment Guidelines by the Work Loss Data Institute (WLDI) are widely used in utilisation review of workers' compensation claims. However, the quality of the ODG does not appear to have been comprehensively evaluated in recent years. In light of new, more rigourous standards for developing guidelines and for performing the systematic reviews that underlie guidelines, this report aims to assess ODG's technical quality (the rigour of development methods) as well as clinical acceptability (the validity of the guideline in the eyes of diverse clinical experts).

To evaluate guideline development methods and supporting systematic reviews, a team of trained appraisers used the AGREE II and modified AMSTAR instruments. ODG's overall score on the AGREE II instrument was 58 per cent (scale: 0–100 per cent), and its overall score on the modified AMSTAR was fair/good (scale: poor to outstanding). Therefore, the development methods, as reported by WLDI, fall short of the highest quality possible, particularly those of the literature reviews supporting the guidelines. The ODG's strengths include an expansive scope, clearly written recommendations, frequent updating, and a well-designed tool for applying recommendations. Important weaknesses include a lack of input from workers with occupational conditions, and limited information about the current process by which ODG chapter development teams identify, select, evaluate, and synthesise evidence. In addition, appraisers were uncertain whether experienced methodologists were involved in development and whether ODG chapter development teams were free of conflicts of interest and had editorial independence from WLDI.

Our AGREE II and AMSTAR appraisals reveal several specific ways in which the technical quality of ODG could be improved during future updates:

1. engaging representatives of workers/labour in guideline development
2. ensuring that experienced methodologists are involved in all aspects of guideline development, particularly in performing literature searches and synthesising evidence
3. improving documentation of search terms, selection criteria, and numbers of studies identified and eligible as well as creating evidence tables that describe published literature for each recommendation
4. documenting and mitigating any conflicts of interest among ODG chapter development teams and ensuring that they have editorial independence so that content is not unduly influenced by WLDI's role as a vendor of utilisation review guidelines and tools.

Following the evaluation of technical quality, a multidisciplinary panel of expert clinicians from the United States and Australia rated the validity of selected guideline content in nine ODG chapters on common occupational conditions. Overall, the panelists found the selected content to

be valid for 41 of 47 topics in these chapters. Panelists judged all of the selected content to be valid in important chapters on low back pain; shoulder injuries; carpal tunnel syndrome; forearm, wrist, and hand injuries; and fitness for duty. For some topics that panelists judged to be of uncertain validity, panelists had less direct expertise. In some cases, panelists questioned whether content was consistent with published evidence or warranted updating, such as diagnostic criteria in the Mental Health and Stress chapter.

Policymakers and payers considering whether to adopt ODG should note that its technical quality has limitations. Nonetheless, clinical experts in diverse fields in two countries considered content related to several important occupational conditions to be valid. As WLDI updates ODG, methodological improvements are warranted to assure that the guideline continues to keep pace with advances in the science of developing guidelines as well as with the expanding base of literature. In this way, applying the ODG in utilisation review procedures is likely to yield the best possible clinical outcomes, in addition to reducing unnecessary healthcare expenditures.

Table of Contents

Tables

Acknowledgements

We would like to thank the following individuals for their contributions to the work: Jennifer Moroney and RAND Australia (based in Canberra), for providing thoughtful comments during the project as well carefully reviewing the final report; Christina Boutsicaris, for her assistance with the guideline technical quality review; our panelists, for their time and thoughtful evaluations of the Official Disability Guidelines chapters; Emily Dao, for her assistance with the ExpertLens™ interface; and Lance Tan, for his assistance with project logistics and the report. The RAND research team would particularly like to acknowledge the contributions of Mark Newton and Chris Colquhoun. Both were exceptionally thoughtful and insightful in the questions they asked and the comments they made. They encouraged the team to consider multiple issues and perspectives, while simultaneously ensuring that we had the editorial independence essential to rigorous, credible policy-related research.

Panelists who participated in the assessment of clinical acceptability include:

- David Crocker, MBBS, occupational medicine, Australia
- Rupali Das, MD, MPH, occupational medicine, United States
- Standiford Helm, MD, MBA, anaesthesia/pain management, United States
- Charles Reitman, MD, orthopaedics/spine, United States
- Michael Fearnside, MBBS, MS, FRACS, neurosurgery/spine, Australia
- Christopher Standaert, MD, physical medicine and rehabilitation, United States
- Julie Fritz, PT, PhD, FAPTA, physical therapy, United States
- Itai Danovitch, MD, MBA, psychiatry, United States.

Abbreviations

AGREE II	Appraisal of Guidelines for Research and Evaluation II
AMSTAR	A Measurement Tool to Assess Systematic Reviews
COI	conflict of interest
CPG	clinical practice guideline
CT	computerised tomography
DSM	Diagnostic and Statistical Manual of Mental Disorders
GDG	guideline development group
MRI	magnetic resonance imaging
NSW	New South Wales
ODG	Official Disability Guidelines
PTSD	posttraumatic stress disorder
U.S.	United States
WLDI	Work Loss Data Institute

1. Introduction

Workers' compensation systems are designed to ensure that workers can get necessary medical care for occupational conditions (Sengupta and Baldwin, 2015). However, some types of care offer limited clinical benefits or may even involve risks that exceed potential benefits. Such care can contribute to high and rising healthcare expenditures. Techniques used by or on behalf of third-party payers to reduce healthcare expenditures by assessing the appropriateness of care provided to individual patients are collectively called utilisation management or utilisation review (Institute of Medicine, 1989). Utilisation review has become widely used in workers' compensation systems in the United States because it can reduce medical costs and ensure appropriate care for injured workers (Wickizer et al., 1999; Wickizer and Lessler, 2002; Wickizer et al., 2004). Insurers and other payers often use guidelines when deciding whether to cover specific treatments for particular patients.

Guidelines that payers apply during utilisation review are slightly different from guidelines that distil research evidence into a more usable form for busy clinicians. The National Academies' definition applies equally well to both types of guidelines:

> Clinical Practice Guidelines are statements that include recommendations intended to optimise patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options (Institute of Medicine, 2011a).

However, the two types of guidelines have different purposes and users, leading to differences in focus and the nature of recommendations. Payers performing utilisation review need to be able to look up specific services to determine whether payment is warranted, whereas clinicians need comprehensive information on how to evaluate, diagnose, and treat specific conditions. In utilisation review, overly restrictive guidelines could prevent patients from accessing potentially beneficial care. Employing rigourously developed and clinically applicable guidelines in utilisation review may enable third-party payers to curtail unnecessary healthcare expenditures while ensuring that patients still have access to the care that helps them achieve optimal outcomes and return to work quickly (Nuckols et al., 2005).

In 2004, the RAND Corporation evaluated five guideline sets for occupational disorders on behalf of the California Department of Industrial Relations. This evaluation considered *technical quality* (rigour of development methods) as well as *clinical acceptability* (validity of the guideline set as determined by diverse clinical experts). One of these guideline sets was the Official Disability Guidelines (ODG®) Medical Treatment Guidelines by the Work Loss Data Institute (WLDI). RAND's study concluded that ODG performed as well or better on technical quality than the other four guideline sets. A multidisciplinary panel of clinical experts ranked the guidelines for clinical acceptability, studying validity and comprehensiveness of guideline

content related to physical therapy; chiropractic manipulation; and surgery for lumbar spine problems, carpal tunnel syndrome, and shoulder injuries. In panelists' rankings, ODG tied for the second-highest–quality guideline in terms of clinical acceptability. The panelists agreed that the ODG was comprehensive and that it was valid for two of the four surgical topics and two of six topics related to physical therapy and chiropractic care. Of note, panelists rated the clinical validity of all five guidelines as uncertain (Nuckols et al., 2005).

In the years since this evaluation was completed, California and many other U.S. states have implemented utilisation review procedures based on guidelines with the aim of improving the appropriateness of care while lowering costs (Texas Department of Insurance, 2014; Colorado Department of Labor and Employment, 2016; Hunt et al., 2016; Pennsylvania Department of Labor and Industry, 2016; State of California Department of Industrial Relations, 2016; Washington State Department of Labor and Industries, 2016). Meanwhile, the number of guidelines that address multiple diverse occupational disorders in workers' compensation settings has declined; the ODG has become the most widely used guideline for utilisation review for occupational disorders. It is used in many U.S. states as well as in other nations (Hunt et al., 2016; Work Loss Data Institute, 2016). A recent study examined the relationship between the adherence to recommendations in ODG and medical care expenditures, and found that patients with adherence scores in the bottom half of the distribution had 13.2 per cent longer duration claims and 37.9 per cent higher medical care costs than patients with adherence scores in the top half of the distribution (Hunt et al., 2016). An investigator on the current research team evaluated ODG's technical quality in 2013 as part of a review of 13 guidelines on the use of opioids for chronic pain. This evaluation found that ODG's technical quality was similar to the average for the 13 guidelines (Nuckols et al., 2014).

Despite widespread adoption rates and evidence for favourable effects on costs, the quality of the ODG has not, to our knowledge, been comprehensively evaluated since the 2004 study (Nuckols et al., 2005). The science of guideline development has advanced since then. In recent years, stricter standards have been issued for guideline development to assure that they are based on the best available evidence, that clinicians view guidelines as valid and acceptable for implementation, and that the implementation of guidelines will lead to improved patient outcomes. A 2011 National Academies report made several strong recommendations for improving the quality of guidelines (see Table 1.1). In addition, in 2011, the National Academies established standards for performing systematic reviews of medical literature; these reviews are fundamental to guideline development. (Due to the length and detail of these recommendations, we have included these standards in the Appendix.)

Given the wide use of ODG and evolving standards for guideline development, a new and independent evaluation of ODG was warranted. This report assesses whether ODG's development methods meet current standards for technical quality and the degree to which clinical experts from diverse specialties find its recommendations to be clinically valid.

Table 1.1. National Academies Recommendations for Guideline Development

1. Establishing transparency	1.1. The processes by which a clinical practice guideline (CPG) is developed and funded should be detailed explicitly and be publicly accessible.
2. Management of conflicts of interest (COIs)	2.1. Prior to selection of the guideline development group (GDG), individuals being considered for membership should declare all interests and activities potentially resulting in COI with development group activity, by written disclosure to those convening the GDG. • Disclosure should reflect all current and planned commercial (including services from which a clinician derives a substantial proportion of income), non-commercial, intellectual, institutional, and patient/public activities pertinent to the potential scope of the CPG. 2.2. Disclosure of COIs within GDG: All COIs of each GDG member should be reported and discussed by the prospective development group prior to the onset of their work. Each panel member should explain how their COI could influence the CPG development process or specific recommendations. 2.3. Divestment: Members of the GDG should divest themselves of financial investments they or their family members have in, and not participate in marketing activities or advisory boards of, entities whose interests could be affected by CPG recommendations. 2.4. Exclusions: Whenever possible, GDG members should not have COIs. In some circumstances, a GDG may not be able to perform its work without members who have COIs, such as relevant clinical specialists who receive a substantial portion of their incomes from services pertinent to the CPG. Members with COIs should represent not more than a minority of the GDG. The chair or co-chairs should not be a person(s) with COI. Funders should have no role in CPG development.
3. Guideline development group composition	3.1. The GDG should be multidisciplinary and balanced, comprising a variety of methodological experts and clinicians, and populations expected to be affected by the CPG. 3.2. Patient and public involvement should be facilitated by including (at least at the time of clinical question formulation and draft CPG review) a current or former patient and a patient advocate or patient/consumer organisation representative in the GDG. 3.3. Strategies to increase effective participation of patient and consumer representatives, including training in appraisal of evidence, should be adopted by GDGs.
4. Clinical practice guideline–systematic review intersection	4.1. CPG developers should use systematic reviews that meet standards set by the Institute of Medicine's Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. 4.2. When systematic reviews are conducted specifically to inform particular guidelines, the GDG and systematic review team should interact regarding the scope, approach, and output of both processes.
5. Establishing evidence foundations for and rating strength of recommendations	5.1. For each recommendation, the following should be provided: • an explanation of the reasoning underlying the recommendation, including a clear description of potential benefits and harms • a summary of relevant available evidence (and evidentiary gaps), description of the quality (including applicability), quantity (including completeness), and consistency of the aggregate available evidence • an explanation of the part played by values, opinion, theory, and clinical experience in deriving the recommendation • a rating of the level of confidence in (certainty regarding) the evidence underpinning the recommendation • a rating of the strength of the recommendation in light of the preceding bullets • a description and explanation of any differences of opinion regarding the recommendation.
6. Articulation of recommendations	6.1. Recommendations should be articulated in a standardised form detailing precisely what the recommended action is and under what circumstances it should be performed. 6.2. Strong recommendations should be worded so that compliance with the recommendation(s) can be evaluated.

7. External review	7.1. External reviewers should comprise a full spectrum of relevant stakeholders, including scientific and clinical experts, organisations (e.g., healthcare, specialty societies), agencies (e.g., federal government), patients, and representatives of the public. 7.2. The authorship of external reviews submitted by individuals and/or organisations should be kept confidential unless that protection has been waived by the reviewer(s). 7.3. The GDG should consider all external reviewer comments and keep a written record of the rationale for modifying or not modifying a CPG in response to reviewers' comments. 7.4. A draft of the CPG at the external review stage or immediately following it (i.e., prior to the final draft) should be made available to the general public for comment. Reasonable notice of impending publication should be provided to interested public stakeholders.
8. Updating	8.1. The CPG publication date, date of pertinent systematic evidence review, and proposed date for future CPG review should be documented in the CPG. 8.2. Literature should be monitored regularly following CPG publication to identify the emergence of new, potentially relevant evidence, and to evaluate the continued validity of the CPG. 8.3. CPGs should be updated when new evidence suggests the need for modification of clinically important recommendations. For example, a CPG should be updated if new evidence shows that a recommended intervention causes previously unknown substantial harm, that a new intervention is significantly superior to a previously recommended intervention from an efficacy or harms perspective, or that a recommendation can be applied to new populations.

SOURCE: Institutes of Medicine, 2011a.

2. Evaluation of Technical Quality

Methods

The National Academy's 2011 standards describe optimal methodological approaches for organisations and individuals engaged in developing guidelines and performing systematic reviews. Adhering to these standards should produce guidelines of high technical quality. These recommended standards do not represent tools for evaluating guidelines and systematic reviews, however. Our evaluation of ODG's technical quality, therefore, used two tools specifically designed for these purposes. First, to assess the overall quality of the guideline development methods, we applied the Appraisal of Guidelines for Research and Evaluation II (AGREE II) Instrument, an international tool designed to assess the quality and reporting of practice guidelines. Second, to examine the critical methods used to assemble the evidence that supports the guideline, we used a previously modified version of A Measurement Tool to Assess Systematic Reviews (AMSTAR), a validated tool for evaluating the methodological quality of systematic reviews (Shea et al., 2009; Nuckols et al., 2014). These tools are generally consistent with the National Academy's recommended standards, as noted below.

AGREE II

The following methods are used to evaluate a guideline according to AGREE II standards (The AGREE Next Steps Consortium, 2009, Brouwers et al., 2010a).

Two to four appraisers rate 23 items across six domains. They also rate the overall quality of the guideline and recommend either for or against use. The six domains are

1. scope and purpose
2. stakeholder involvement
3. rigour of development
4. clarity of presentation
5. applicability
6. editorial independence (Brouwers et al., 2010a; Brouwers et al., 2010b).

Overall, the six domains and 23 items are consistent with the 2011 recommendation standards for guideline development by the National Academies (Institute of Medicine, 2011a).

Appraisers rate each item on a scale from one (missing information or inadequate documentation) to seven (excellent documentation of all information relevant to the AGREE II item). Appraisers generally conduct their reviews independently and are not expected to reach agreement. After all appraisers complete their evaluations, the AGREE II ratings are summed across all appraisers, and scaled domain scores (0 to 100 per cent) are calculated based on the

difference between the maximum and minimum possible scores (Brouwers et al., 2010a; Brouwers et al., 2010c).

AMSTAR

The original AMSTAR instrument contains 11 questions related to methods used in performing systematic reviews. These standards are consistent with, but not as strict as, the recommended standards for performing systematic reviews by the National Academies (Institute of Medicine, 2011b). In prior work, we modified the instrument to facilitate reliable scoring by dividing each of the 11 questions into one to seven subquestions (Nuckols et al., 2014). Appraisers rate subquestion(s) within each domain (yes, no, cannot answer, or not applicable), the 11 main domains (poor, fair, good, excellent, or outstanding), and the overall quality of the review (poor, fair, good, excellent, or outstanding).

Our appraisal team for both AGREE II and AMSTAR ratings consisted of three clinicians and one research associate with training in guideline evaluation. Appraisals were based on content posted on the ODG website during April 2016, as well as two interviews with individuals involved in developing ODG. Appraisers focused on the overall literature search methodology reported; they did not evaluate the quality of any systematic reviews cited by ODG. WLDI reported the literature review methods that its chapter development teams tend to use, but it did not provide separate information for each individual chapter or report the methods actually used. For the AGREE II instrument, RAND appraisers discussed their ratings as a group and were given the opportunity to change their ratings based on the discussion; as noted above, scoring the AGREE II instrument does not require reviewers to reach consensus. For the AMSTAR instrument, appraisers conducted similar discussions and then reached agreement on scores for each question, each domain, and for the overall score.

After initial ratings were complete, WLDI informed RAND that the guidelines had been updated with additional details regarding methodology, including details of the internal quality assurance processes and search strategies for the systematic review (and other systematic review methodologies) supporting the guidelines. In addition, WLDI added an appendix that specifically addressed methodological details requested under AMSTAR, although it is unclear whether this reflects a change in how reviews will be conducted in the future, or reflects a change in describing how reviews were performed in the past. RAND appraisers met to discuss these updates in August 2016 and re-rated all 11 AMSTAR domains, as well as the Rigour of Development section and overall rating of the AGREE II instrument. The results of the appraisals changed only modestly, and the current report includes the results of the final evaluation.

Results

Guideline Development: AGREE II Instrument

Table 2.1 summarises the AGREE II scores. The overall score was 58 per cent, while individual domain scores ranged from 55 per cent to 75 per cent. All four appraisers recommended the guideline for use, ideally with future modifications to technical quality. In particular, appraisers had concerns about fundamental questions covered in the Rigour of Development domain. Table 2.2 lists strengths, limitations, and major recommendations for improvement.

Table 2.1. AGREE II Appraisal

Domain	Score
Scope and Purpose	64%
Stakeholder Involvement	67%
Rigour of Development	55%
Clarity of Presentation	75%
Applicability	74%
Editorial Independence	69%
Overall	58%
Recommended for Use	Yes, with modifications

Table 2.2. Details on AGREE II Ratings, Strengths, Limitations, and Recommendations for Improvement

Domains and Questions	Median Score (Range 1–9)	Strengths	Limitations	Major Recommendations for Improvement
Scope and Purpose				
1. The overall objective(s) of the guideline is (are) specifically described.	5.5 (5–6)	The ODG mission statement in Appendix B clearly articulates the guideline objectives.	AGREE II stipulates that this information should ideally be in the Background section.	
2. The health question(s) covered by the guideline is (are) specifically described.	4 (4–5)	Attempts to show all possible interventions and treatments for each condition; interventions are usually specific.	Guidelines describe evidence for particular interventions, but in some places it does not clearly state the specific conditions to which it applies.	
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	4.5 (3–5)	Guideline is intended to cover treatments for all work-related disorders within working-age population.	The guideline does not provide clinical case definitions, which means that treatment decisions could be based on incorrect diagnoses.	
Stakeholder Involvement				
4. The guideline development group includes individuals from all relevant professional groups.	5 (5–6)	The "Editorial Advisory Board" includes a wide range of relevant clinical disciplines and persons engaged in utilisation review.	Methodologists' experience and role in reviewing overall methods is not well documented.	Ensure that a qualified methodologist is involved in all aspects of guideline development, particularly performing the literature searches and syntheses.

Domains and Questions	Median Score (Range 1–9)	Strengths	Limitations	Major Recommendations for Improvement
5. The views and preferences of the target population (patients, public, etc.) have been sought.	2.5 (2–3)	Guideline has clear rules for accepting comments from public, workers, and other stakeholders.	Although comments on guidelines are open, workers and patients may not know how to access guidelines, which limits their ability to provide input.	Engage labour or patient advocacy groups in guideline development or review prior to release.
6. The target users of the guideline are clearly defined.	6 (5–6)	The ODG is responsive to the needs of the primary users, claims managers.		
Rigour of Development				
7. Systematic methods were used to search for evidence.	3 (2–4)	Primary databases and timeframe for searches are documented.	Guideline includes limited information on the search strategies used and how results were synthesised. Evidence was described in relation to broad topics (e.g., shoulder injuries), not specific conditions (e.g., rotator cuff tears).	As updates are made to ODG, improve documentation of search terms, selection criteria, and numbers of studies identified and eligible.
8. The criteria for selecting the evidence are clearly described.	3 (2–3)	The guideline seeks to include a broad range of available evidence, which is limited for many topics.	Guideline includes limited information on study inclusion/exclusion criteria, methods for evaluating study quality, reviewer training, reconciliation processes, and extraction of data from the articles. No evidence tables were included summarising eligible studies.	Describe literature search methods in detail.
9. The strengths and limitations of the body of evidence are clearly described.	2 (2–3)	Study designs included in the review are clear. Each study is ranked and the ranking is provided in the Reference Summaries.	Evidence summaries do not include information on bias in primary studies. Most assessments of evidence for particular interventions do not provide overall assessment (or grade) to entire body of evidence, although individual studies are often graded.	Include more information in the Procedure Summaries about the quality of the evidence summarised.
10. The methods for formulating the recommendations are clearly described.	5 (5–5)	The process for reviewing the evidence and updating recommendations is described in broad terms. ODG chapter leads and teams appear to identify the studies, extract information, and formulate recommendations for Advisory Board to vote on.	ODG chapter leads appear to have a substantial influence on the guideline content. Few details are provided on the methods of formulating and voting on recommendations, such as how ODG ensures that board members vote yes or no on every treatment.	Include a broader number of individuals in the process of formatting recommendations. Describe the consensus process in greater detail in public documents.
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	5.5 (5–6)	Recommendations generally consider both harms and benefits.	Harms and benefits do not appear to have been extracted systematically. Guidelines do not consistently describe the magnitude of effects.	

Domains and Questions	Median Score (Range 1–9)	Strengths	Limitations	Major Recommendations for Improvement
12. There is an explicit link between the recommendations and the supporting evidence.	5 (4–6)	The guidelines generally tie recommendations for individual treatments to citations of relevant evidence.	It is not clear whether the full body of evidence is considered in each recommendation.	Link narrative summaries to evidence tables that at least include study designs, sample sizes, major findings, and potential biases.
13. The guideline has been externally reviewed by experts prior to its publication.	5 (4-6)	The guideline obtains input from specialty societies and other sources, as well as an advisory board. A large number of diverse professionals are involved.	The Advisory Board both votes on the recommendations and participates in the “external” reviews. When and how specialty societies and other stakeholder input is used is unclear.	
14. A procedure for updating the guideline is provided.	6 (6-6)	The guideline is updated at least every six months. The updating process includes the addition of new technologies not found in the database. Updates are available for users to view.	The details of how evidence is weighed for new updates is unclear.	
Clarity of Presentation				
15. The recommendations are specific and unambiguous.	6 (6–6)	Recommendations are generally specific.	Interventions and conditions are not always fully defined. For example, for "manipulation" for shoulder disorders, the various manipulation approaches are not provided.	
16. The different options for management of the condition or health issue are clearly presented.	4.5 (3–5)	The different options are described and criteria for switching from one option to another are sometimes included, particularly under the "Treatment Planning" chapter.	The quality of treatment planning descriptions varies across topics. Treatment options for patients with less common conditions are not always included.	
17. Key recommendations are easily identifiable.	6 (6–7)	Recommendations in UR Advisor app are generally unambiguous yes/no statements that are easy to find.	Determining which recommendations are higher vs. lower priority is hard to discern. The website design makes it difficult to find some items.	
Applicability				
18. The guideline describes facilitators and barriers to its application.	6 (3–7)	Appendix presents the barriers and facilitators to application. From the perspective of a utilisation review claims manager, the guideline appears to have few barriers to application.	The material is less accessible to clinicians and patients/workers.	
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	7 (6–7)	Automatic approval codes are helpful for linking recommendations to practice. Claims managers should have little difficulty applying	Little information is provided on what to do when it is uncertain whether patients have the condition in question.	

Domains and Questions	Median Score (Range 1–9)	Strengths	Limitations	Major Recommendations for Improvement
		ODG.		
20. The potential resource implications of applying the recommendations have been considered.	5 (1–6)	The purpose of the guideline is for utilisation review, so it is designed to reduce resource use.	The guidelines themselves do not mention cost considerations.	
21. The guideline presents monitoring and/or auditing criteria.	5 (4–7)	The automatic approval codes are monitoring criteria. The UR Adviser app includes functions to audit care consistency with ODG. This is far more developed than many guidelines.	Assessing usage frequency with page hits on the UR Advisor app does not equate to auditing how the guideline is being followed.	
Editorial Independence				
22. The views of the funding body have not influenced the content of the guideline.	5.5 (1–7)	A large group of individuals with diverse qualifications is involved in reviewing the guideline.	There is no statement about the editorial independence or employment status of the chapter leads or Advisory Board members. Guideline materials assert that the "funders" are individual subscribers. However, individual subscribers' decisions are often driven by state policy decisions.	Ensure guideline chapter leads have editorial independence. Provide greater transparency regarding Advisory Board compensation.
23. Competing interests of guideline development group members have been recorded and addressed.	5.5 (5–6)	The authors state that competing interests are documented, and an attempt is made to get evaluations outside those biased opinions or to counterbalance with an opposing bias.	A list of competing interests is not provided.	
Overall Score				
Overall quality of guideline.	4 (4–5)			

ODG's strongest domain scores were in Clarity of Presentation (75 per cent) and Applicability (74 per cent). The guideline provides very specific recommendations and includes a utilisation review tool for tying recommendations to clinical questions. However, the guideline provides limited guidance on eligibility for the recommendations, including clinical criteria or case definitions, and when clinical circumstances warrant considering particular interventions. Nonetheless, if a worker has already been diagnosed with a particular condition, the utilisation review tool provides very specific recommendations regarding interventions. Published clinical guidelines often score much lower than ODG does on applicability (Nuckols et al., 2014).

Scope and Purpose received a score of 64 per cent. The ODG states its mission most clearly in a methodological appendix geared toward addressing the AGREE II instrument. According to that appendix, ODG aims to provide "best practices (as supported by evidence) to improve workplace health (including functional status), reduce injury claims, and improve rates of return

to work." However, the AGREE II instrument specifically recommends that the opening paragraphs in a guideline describe its scope and purpose and recommends that such information should be easy to find. Information related to scope and purpose was more limited in the main background section.

Stakeholder Involvement received a score of 67 per cent. ODG describes its target audience clearly: "independent treating physicians, allied healthcare providers, medical review organisations, insurance claims professionals, nurse case managers, managed care organisations, and regulatory authorities." ODG's advisory board includes clinical experts from many relevant disciplines. In addition, WLDI has performed literature reviews that consider patient outcomes. Most notably, patients/workers are able to submit suggested changes to WLDI, a desirable feature that many guidelines lack. However, it is unclear how patients and workers would know that their care is subject to ODG utilisation review, how to access the ODG, or how to submit comments to WLDI. Furthermore, there is little evidence that WLDI has a mechanism for consulting directly with patients or workers to obtain meaningful input on most topics. This falls short of the Institute of Medicine's recommendation to include patients and/or patient advocates on guideline development teams, as well as the AGREE II recommendation to consult or interview patients and to perform literature reviews that specifically address patients' values and preferences (Institute of Medicine, 2011a).

As part of an assessment of stakeholder involvement, AGREE II also specifically recommends that appraisers evaluate whether the review group includes at least one methodology expert. AGREE II notes that an epidemiologist, statistician, and library scientist should be involved. Evidence of these sorts of skills could include having a graduate degree in these or related fields, publishing systematic reviews in peer-reviewed medical journals, or helping to lead the development of prior peer-reviewed guidelines. WLDI introduced RAND to chapter leads who had highly relevant clinical expertise but who lacked evidence of substantial methodological expertise related to performing literature searches or developing guidelines. Subsequently, WLDI revised the guideline to state that two methodologists are involved in development, but included no information on the names, titles, roles in development of the guideline, or relevant expertise. It was not clear to the RAND evaluation team whether these individuals were involved in developing the existing guideline or newly recruited to guide future updates. Scores for this domain were therefore reduced due to the lack of patient/worker involvement and the uncertainty about the involvement of methodological experts.

The Editorial Independence domain received a score of 69 per cent. AGREE II recommends that "competing interests of guideline development group members have been recorded and addressed." WLDI reports obtaining information from Advisory Board members on conflicts of interest. However, these potential conflicts of interest are not described in detail in the ODG documentation. AGREE II also recommends that "the views of the funding body have not influenced the content of the guideline." WLDI pays chapter leads and other team members involved in ODG development, but the guideline does not explain how such persons may have

been assured editorial independence from WLDI or ODG subscribers (typically, insurance companies and state workers' compensation agencies).

Rigour of Development received the lowest domain score (55 per cent). ODG is updated continuously, which is a major strength, especially in comparison to typical practices for many systematic reviews and guidelines (Shekelle et al., 2001; Shojania et al., 2007). ODG relies, where possible, on systematic reviews performed by other groups because systematic reviews tend to be high-quality sources of evidence. However, ODG is designed to include all treatments that may be requested in workers' compensation settings, and systematic reviews do not exist for many topics. When systematic reviews are not available, WLDI conducts its own reviews of the primary literature. AGREE II stipulates that specific information should be provided on the literature review methods used during guideline development. Although WLDI describes the reviews as "comprehensive and ongoing," they lacked detail on search terms, inclusion and exclusion criteria for specific topics, and how evidence was extracted, tabulated, and summarised. Furthermore, according to AGREE II, a guideline should clearly state when a "recommendation is informed primarily by consensus of opinion by the guideline group, rather than the evidence." ODG does not explicitly note this limitation in many instances when a comprehensive systematic review is unavailable. These are basic standards for systematic reviews performed during guideline development. The lack of information on these fundamental methodological details may reflect limited involvement by individuals with methodological expertise, as discussed above. (See below under the AMSTAR Instrument for more details on systematic review methods.)

Systematic Reviews: Modified AMSTAR Instrument

Table 2.3 presents results for the modified AMSTAR appraisal. Appraisers agreed that the overall quality of ODG literature reviews was fair to good, based on the documentation available in the ODG guideline, as well as interviews with ODG developers.

Table 2.3. Modified AMSTAR Appraisal

Domains and Questions	Group Rating*
1. Was an *a priori* design provided?	Good
2. Was there duplicate study selection and data extraction?	Fair
3. Was a comprehensive literature search performed?	Fair
4. Was the status of publication (i.e., grey literature) used as an inclusion criterion?	Outstanding
5. Was a list of studies (included and excluded) provided?	Good
6. Were the characteristics of the included studies provided?	Fair
7. Was the scientific quality of the included studies assessed and documented?	Good
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	Fair
9. Were the methods used to combine the findings of studies appropriate?	Good
10. Was the likelihood of publication bias assessed?	Fair
11. Were any conflicts of interest stated?	Fair
Overall Rating	Fair to Good

The appraisers noted that the ODG literature reviews have an expansive scope and are updated very frequently. In addition, WLDI appears to use some standard methods for systematic reviews, including conducting broad searches of the medical literature, including grey literature; having at least two reviewers assess the eligibility of the articles identified; assessing the scientific quality of the eligible articles; and summarising findings of at least some studies. However, WLDI only provides a broad description of how they perform searches and does not provide information on how searches were actually conducted for each chapter. Search strategies are incompletely described, and there is little documentation of standard processes for evidence extraction, tabulation, and synthesis. Furthermore, authors do not clearly report how evidence was synthesised, and the methods for choosing which evidence to incorporate in the guideline appear to involve a high degree of subjectivity.
Table 2.4 provides further details and comments.

Table 2.4. Details of Modified AMSTAR Appraisal

AMSTAR Questions and Subquestions	Group Consensus	Comments
1. Was an *a priori* design provided? The research question and inclusion criteria should be established before the review is conducted.	**Good**	
The research question was established before the review was conducted.	Yes	
The inclusion criteria were established before the review was conducted.	No	The inclusion criteria appear to be determined after the results of the literature search are known (e.g., if a prior systematic review exists, other evidence is not considered).
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	**Fair**	The explanations of study selection and data extraction provided very little detail and did not include methods for each chapter in ODG. This is a major concern due to the possibility of bias in the choice of studies included in the narrative summaries of the evidence.
Study Selection: There were at least two independent reviewers for study selection.	Yes	
Study Selection: There was a consensus procedure for disagreements about study selection.	Uncertain	Details of this process were unclear.
Data Extraction: There were at least two independent reviewers for extracting data from individual studies.	Yes	
Data Extraction: There was a consensus procedure for disagreements about data extracted from individual studies.	Uncertain	The data extraction process appears to overlap with the formulation of recommendations in terms of decisions to either include or exclude studies.

3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g., CENTRAL, Embase, Medline). Keywords and/or medical subject Heading terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialised registers, or experts in the particular field of study, and by reviewing the references in the studies found.	**Fair**	
At least two electronic sources were listed (e.g., CENTRAL, Embase, and Medline).	Yes	
The document listed the years searched.	Yes	
Keywords and/or medical subject heading terms must be stated.	No	Only sample terms are provided, and the sample terms include only one word or phrase per clinical question or treatment being considered. High-quality searches should generally involve a large list of synonyms for each clinical question.
Searches were supplemented by reviewing the references in the studies found.	Yes	
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language, etc.	**Outstanding**	
Types of publications: The authors described the types of publications that were included and/or excluded from the analysis.	Yes	
Grey literature: The analysis should include high-quality grey literature.	Yes	
	Yes	
Language: The authors described the languages that were included and/or excluded from the analysis.		
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	**Good**	
A list of included studies was provided.	Yes	
A list of excluded studies was provided.	No	(Note: the appraisal team has yet to find a guideline with a systematic review that does this.)

6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions, and outcomes. The ranges of characteristics in all of the studies analysed (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported.	**Fair**	
Data from original studies was provided in aggregated form, such as a table or narrative summary.	Yes	Narrative summaries are provided.
The aggregated data included information on participants (clinical condition(s) studied), interventions, and outcomes (severity of pain).	Uncertain	The narrative summaries are brief. No detail is provided on the participants, interventions, and outcome measures used.
7. Was the scientific quality of the included studies assessed and documented? *A priori* methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomised, double-blind, placebo-controlled studies, or allocation concealment as inclusion criteria); for other types of studies, alternative items will be relevant.	**Good**	Although WLDI states that the scientific quality of the evidence was assessed, no documentation is provided on the results of those assessments.
Was the scientific quality of the individual studies assessed?	Yes	
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigour and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	**Fair**	
For each individual recommendation, was the quality of the evidence supporting the recommendation graded?	Uncertain	No information was provided on the quality of the evidence supporting each recommendation.
Did the study use a standard, widely accepted grading scale?	No	The grading scale cited is not widely used in the performance of systematic reviews or guideline development.
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, I^2). If heterogeneity exists, a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).	**Good**	WLDI states that these methods are used, but no details are provided on how findings were synthesised for individual topics. The reader does not know whether these methods were applied to all clinical topics included or how they were applied.
Were quantitative methods used to combine findings?	No	
If quantitative methods were used, was homogeneity assessed?	N/A	
Were qualitative methods used to combine	Yes	

findings?	
If qualitative methods were used, did they assess the following?	
Risk of bias	Yes
Consistency	Yes
Directness	Yes
Precision	Yes
Dose response	Uncertain
Confounding	Yes
Strength of association	Yes
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger's regression test).	**Fair**
Was a meta-analysis or similar quantitative method used to combine results?	No
Was a formal assessment of publication bias performed?	No
Did it include graphical aids (e.g., funnel plot)?	No
Did it include statistical tests (e.g., Egger's regression test)?	No
If no formal assessment of publication bias was performed, did the study comment on the likelihood of publication bias based on grey literature, searches on clinical trials registries, or other sources?	Uncertain
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and in the included studies.	**Fair**
Were potential sources of support for the review acknowledged?	Yes
Were potential sources of support for the included studies provided?	Uncertain

As noted above, ODG uses, where possible, systematic reviews published by prior groups, including the Cochrane Collaboration. Systematic reviews are generally considered to be high-quality evidence. However, systematic reviews do not exist for many topics, particularly given the extremely wide range of topics covered by ODG. Furthermore, the quality of systematic reviews can vary, and ODG does not formally rate the quality of the systematic reviews it uses, nor indicate whether such reviews are up-to-date. When systematic reviews are lacking for a subject, ODG appears to use other types of literature that it identifies through its own searches, but critical information is lacking on how such literature is identified and selected for inclusion.

It is unclear whether the evidence summaries include all eligible articles or only articles that the chapter development teams preferred to include.

3. Evaluation of Clinical Acceptability

Methods

In prior work on guidelines for occupational conditions, we determined that rigourous development methods (technical quality) did not assure that expert clinicians found these guidelines to be valid (clinical quality) (Nuckols et al., 2008). Clinicians are important stakeholders in workers' compensation systems, and they are responsible for advocating for their patients. Moreover, expert clinicians are in the best position to identify the flaws in a guideline, such as disagreements with published evidence or outdated information. Finally, guidelines that have high clinical quality are more likely to yield the best possible clinical outcomes for patients. Therefore, we assessed whether experts in the care of occupational disorders found these guidelines clinically acceptable.

Panel Recruitment and Characteristics

We recruited panelists from a wide array of backgrounds and viewpoints to reflect the variety of specialties involved in treating work-related conditions. Potential panelists were identified based on expertise, participation in prior RAND panels, or the development of Australian guidelines on musculoskeletal disorders or pain management. We also asked key contacts and organisations for recommendations. Potential panelists were contacted about participation and asked to send their most recent curriculum vitae and respond to a questionnaire about their experiences treating occupation-related conditions and serving on expert panels. Initially, ten panelists were selected; however, two panelists were unable to participate.

In October 2016, this multidisciplinary panel rated the validity of the ODG clinical content. The final panel represented relevant clinical disciplines, including pain management, occupational medicine, physical therapy, orthopaedics, neurosurgery, psychiatry, and physical medicine and rehabilitation. Panelists were from diverse practice settings; two panelists were from Australia and six were from the United States. Additionally, panelists practiced in a range of settings; three practiced in academic settings and five in community settings. To avoid conflicts of interest, we asked potential panelists to report current or prior financial ties to ODG or the WLDI; none reported having such ties.

Rating Process

Panelists evaluated the validity of ODG clinical content in a three-round, modified-Delphi process. Panelists shared their perceptions of the guideline content via the RAND ExpertLens™, an online Delphi platform for expert elicitation and stakeholder engagement. During Round One, panelists reviewed the ODG guideline and rated the validity of content for certain aspects of care

related to several broad clinical domains. Additionally, each panelist was assigned a specific topic and drafted a brief summary (four to five short paragraphs) on the strengths and limitations of recommendations related to that clinical domain. During Round Two, the RAND team led an asynchronous online discussion of each selected topic for all relevant aspects of care, including Round One ratings and summaries. Finally, in Round Three, panelists were given the opportunity to revise their original ratings in response to the discussion and in consideration of ratings by other panelists.

The panelists rated nine guideline chapters. These chapters mostly covered common conditions that contribute to substantial healthcare utilisation:

1. low back pain
2. shoulder injuries
3. carpal tunnel syndrome
4. neck and upper back injuries
5. forearm, wrist, and hand injuries
6. hip and pelvis injuries
7. mental illness and stress
8. pain (panelists were instructed to focus on chronic pain)
9. fitness for duty assessment.

Within each of the chapters on musculoskeletal conditions and pain, panelists were asked to rate content related to diagnostic imaging, physical therapy, medication therapy, chiropractic care, nonoperative procedures, and surgery. For mental illness and stress, panelists rated psychological therapy (chiropractic care and surgery did not apply). For fitness for duty, panelists rated content related to return-to-work planning.

Panelists were also asked to rate sections (medication therapy, physical therapy, etc.) within each remaining chapter, as well as within the guideline as a whole. However, several panelists did not submit such ratings because they found it difficult to provide overall scores that summarised all of ODG's clinical chapters or the remaining ten chapters. We, therefore, did not analyse panelist assessments of the overall guideline and of the remaining ten chapters.

Rating Criteria

Panelists rated the validity of ODG recommendations using prespecified rating criteria. The guidelines related to *providing* or *covering* each selected *aspect of care* for each selected *type of condition.* In addition, panelists rated the remaining content and the guideline as a whole.

We instructed panelists to rate guideline recommendations as *valid* if they met the following criteria:

- evidence-based, or consistent with current, widely held expert opinion (recommendations that are both evidence-based and consistent with current, widely-held expert opinion being assigned the highest validity scores)
- appropriate (they offer patients a favourable balance of potential benefits and risks)

- preferred (no alternative approach to care is substantially more appropriate; the guideline is evidence-based or consistent with current, widely held expert opinion).

Panelists rated guideline content on a nine-point scale, with nine as the highest rating. Ratings were interpreted as follows:

- one to three: guideline recommendations are not valid
- four to six: guideline recommendations are of uncertain, intermediate, or mixed validity
- seven to nine: guideline recommendations are valid.

During the analysis, ExpertLens provided ratings tabulated automatically using the RAND/UCLA Appropriateness Method (Fitch et al., 2001; Khodyakov et al., 2016). We used the three categories listed below.

1. not valid: a median rating of one to three without disagreement
2. uncertain validity: a median rating of four to six, or any median with disagreement
3. valid: a median rating of seven to nine without disagreement.

Disagreement meant that the median was at one end of the range (e.g., seven to nine) and two or more panelists rated at the opposite end of the range (e.g., one to three).

Results

Overall, the panelists found the selected content in the chapters to be valid; they judged some specific content to be of uncertain validity. Panelists rated the content in each clinical area (medication therapy, diagnostic imaging, etc.) of the following topics to be valid: low back pain; shoulder injuries; carpal tunnel syndrome; forearm, wrist, and hand injuries; and fitness for duty. For the remaining topics, panelists judged most of the selected content as valid. Content of uncertain validity related to nonoperative procedures for neck and upper back injuries; chiropractic care for hip injuries and chronic pain; medication therapy for neck and upper back injuries; surgery for chronic pain; and diagnostic imaging for mental illness and stress. Panelists noted that some content extended beyond the published evidence or warranted updating. Table 3.1 summarises the ratings; we later include selected critiques by the panelists.

Table 3.1. Summary of Clinical Acceptability Ratings

	Low Back Pain	Shoulder	Carpal Tunnel	Neck and Upper Back	Forearm, Wrist, Hand	Hip	Chronic Pain	Mental Illness and Stress	Fitness for Duty
Diagnostic Imaging	valid	valid	valid	valid	valid	valid	valid	uncertain	—
Physical Therapy	valid	valid	valid	valid	valid	valid	valid	—	—
Psychological Therapy	—	—	—	—	—	—	—	valid	—

Medication Therapy	valid	valid	valid	uncertain	valid	valid	valid	valid	—
Chiropractic Care	valid	valid	valid	valid	valid	uncertain	uncertain	—	—
Nonoperative Procedures	valid	valid	valid	uncertain	valid	valid	valid	valid	—
Surgery	valid	valid	valid	valid	valid	valid	uncertain	—	—
Return to Work	—	—	—	—	—	—	—	—	valid

Low Back Pain

The guidelines for low back pain were rated by seven or eight panelists (see Table 3.2). All categories of management options rated as valid.

Table 3.2. Ratings for Low Back Pain Interventions

Low Back Pain	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	7	Valid	8	4–8
Physical Therapy	7.5	Valid	8	4–8
Medication Therapy	8	Valid	8	6–8
Chiropractic Care	7	Valid	8	3–8
Nonoperative Procedures	7	Valid	7	6–8
Surgery	7.5	Valid	8	5–8

Diagnostic Imaging for Low Back Pain

Panelists rated the guidelines for diagnostic imaging of low back pain as valid, with a median rating of seven. Panelists commented that the recommendations were generally appropriate, with one panelist commenting that the guidelines appropriately cover the general approach to diagnostic imaging, including patient history and potential red flags. However, several panelists identified outdated literature, which, in some instances, led to recommendations that panelists felt were inconsistent with current evidence-based practices. Key areas of concern highlighted by the panelists included recommendations for discography, radiographs in the absence of pain, and oblique X-rays for spondylolysis in adults. For intraoperative monitoring, the language seems to imply that monitoring is at the discretion of the surgeon and fails to mention clear indications for monitoring.

Physical Therapy for Low Back Pain

The panel rated the guidelines for physical therapy for low back pain as valid, with a median rating of 7.5. A panelist commented that the guidelines provided well-documented, evidence-based explanations of the efficacy of physical therapy for low back pain. However, several

panelists also suggested that recommendations regarding the specific quantities of treatment sessions were not evidence-based.

Medication Therapy for Low Back Pain

Panelists agreed that the guidelines on medication therapy for low back pain were valid and generally appropriate, with a median rating of eight. One panelist commented that the guideline narrative exhibited a high degree of validity despite considerable variation in the quality of the evidence. One panelist also pointed out that concerns about certain medications (e.g., opiates) are well-outlined. However, an Australian panelist noted that some of the medications listed had unfamiliar names. Additionally, panelists expressed concern that some recommendations supported several interventions that had limited or no evidence, such as using antibiotics or oral steroids for back pain and using opioids or muscle relaxers for acute, severe cases.

Chiropractic Care for Low Back Pain

Panelists agreed that the guidelines on chiropractic care for low back pain were valid, with a median rating of seven. Two panelists commented that the recommendations were reasonable and well done. Two panelists pointed out that chiropractic care differs in Australia as compared to the United States (with respect to the initial provider and provider skills). Two panelists commented that ODG appeared to incorrectly equate chiropractic care with any form of manipulation. One of these panelists commented that chiropractors do more than just manipulation, while another panelist pointed out that much of the evidence cited for manipulation was for physical therapy manipulation, not for chiropractic manipulation.

Nonoperative Procedures for Low Back Pain

Panelists agreed that the guidelines addressing nonoperative procedures for low back pain were valid, with a median rating of seven. Panelists commented that the guidelines were comprehensive and appropriately discussed the lack of evidence or variability of evidence for certain procedures. The recommendations for acupuncture, bracing, and powered traction devices were considered reasonable. However, panelists had differing opinions on recommendations for facet and epidural steroid injections, with some panelists approving of use and even suggesting an expansion of use, and others suggesting that the recommendations were not consistent with current evidence.

Surgery for Low Back Pain

Panelists agreed that the surgical recommendations for low back pain were valid, with a median rating of 7.5. Overall, panelists commented that the guidelines effectively covered the broad range of relevant operations and included well-referenced evidence. The recommendations for decompression; discectomy; and newer products, such as flexible fusions, were considered reasonable. One panelist considered the recommendations on bone-morphogenic protein and fusion for all degenerative disc disease to be too rigid. Additionally, this panelist critiqued

ODG's recommendations against surgical treatments for degenerative disc disease and degenerative scoliosis.

Shoulder Injuries

Seven to eight panelists provided ratings for the appropriateness of guidelines related to shoulder injuries. All management categories were rated as valid (see Table 3.3).

Table 3.3. Ratings for Guidelines for Shoulder Injuries

Shoulder	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	8	Valid	7	7–9
Physical Therapy	7	Valid	8	7–8
Medication Therapy	7	Valid	8	5–8
Chiropractic Care	7	Valid	8	6–8
Nonoperative Procedures	7	Valid	8	5–7
Surgery	7.5	Valid	8	6–9

Diagnostic Imaging for Shoulder Injuries

Panelists agreed that the guidelines on diagnostic imaging for shoulder injuries were valid with a median rating of eight. One panelist questioned the recommendation for plain film prior to other scans, but thought that the magnetic resonance imaging (MRI) and computerised tomography (CT) recommendations were reasonable. Another panelist expressed concerns about ODG recommending CT scans for labral tears instead of MRI.

Physical Therapy for Shoulder Injuries

Panelists agreed that the guidelines on physical therapy for shoulder injuries were valid, with a median rating of seven. Overall, panelists found this section to be generally reasonable, especially given the challenge of describing such a wide range of conditions. One panelist commented that the guideline's limits on the number of physical therapy sessions were generally acceptable, but that they did not appear to be evidence-based, and the guideline did not include a citation indicating the source of these limits. Another panelist agreed with these points, expressing concern about recommending limits without evidence.

Medication Therapy for Shoulder Injuries

Panelists agreed that the guidelines on medication therapy for shoulder injuries were valid, with a median rating of seven. Generally, panelists noted that this section primarily referenced the pain chapter and focused more on injections. One panelist particularly appreciated the recommendation of pulsed radiofrequency of the suprascapular nerve. However, another panelist noted that the summaries for nonsteroidal anti-inflammatory drugs (NSAIDs) were not supported

by evidence. Panelists also suggested that ODG add precautions regarding the use of oral steroids and recommend against prescribing opioids for acute shoulder pain, even if severe.

Chiropractic Care for Shoulder Injuries

Panelists agreed that the guidelines on chiropractic care for shoulder injuries were valid, with a median rating of seven. Panelists considered these recommendations to be generally reasonable given the limited role of chiropractic care for these injuries, although they noted that the section conflated manual and manipulation therapy. One expert from Australia noted that chiropractic care practice was much different in Australia than in the United States.

Nonoperative Procedures for Shoulder Injuries

Panelists agreed that the guidelines on nonoperative procedures for shoulder injuries were valid, with a median rating of seven. Generally, the recommendations were considered adequate, with one panelist appreciating the recommendations against electrical stimulation, dry needling, and other passive modalities. Another panelist agreed with ODG's recommendation for steroid injections and suprascapular nerve injections in some cases. However, panelists also disagreed with the guidelines in several cases. One panelist expressed concerns about the lack of evidence supporting the recommendations for platelet-rich plasma and pulsed radiofrequency. Another panelist noted some concerns about the steroid recommendations, especially if bursitis or calcific tendinitis were present on imaging.

Surgery for Shoulder Injuries

Panelists agreed that the guidelines on surgery for shoulder injuries were valid, with a median rating of 7.5. Generally, panelists considered the guidelines for this section to be appropriate but added a few suggestions: One panelist suggested adding information about the risk of adhesive capsulitis after surgical interventions (e.g., acromioplasty); another panelist questioned the timeframes for acute, subacute, and chronic surgical interventions.

Carpal Tunnel Syndrome

Eight panelists provided ratings for the appropriateness of guidelines related to carpal tunnel syndrome. All categories of management options were rated as valid (see Table 3.4).

Table 3.4. Ratings for Guidelines on Carpal Tunnel Syndrome

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	8	Valid	8	6–9
Physical Therapy	7	Valid	8	3–8
Medication Therapy	7	Valid	8	6–9
Chiropractic Care	8	Valid	8	4–9
Nonoperative Procedures	8	Valid	8	4–8
Surgery	8	Valid	8	7–9

Diagnostic Imaging for Carpal Tunnel Syndrome

Panelists agreed that the recommendations for diagnostic imaging for carpal tunnel syndrome were valid, with a median rating of eight. Panelists commented that the evidence for the recommendations was well-referenced and the timelines and indications seemed appropriate. The panelists agreed with the ODG recommendations for limited use of ultrasound or MRI for instances where the diagnosis is unclear. However, one panelist thought that ODG should note that MRI is indicated for mass lesions; another believed that it may be useful to assess the size of the canal in uncertain diagnoses.

Physical Therapy for Carpal Tunnel Syndrome

Panelists agreed that the guidelines for physical therapy for carpal tunnel syndrome were valid, with a median rating of seven. Panelists agreed that the evidence was well-referenced and the guidelines were adequate. One panelist thought that the presentation of the range of therapies was useful.

Medication Therapy for Carpal Tunnel Syndrome

Panelists agreed that the ratings for medication therapy for carpal tunnel syndrome were valid, with a median rating of seven. One panelist commented that this section was limited but appropriately reflected the available literature. Most panelists were in agreement with ODG's recommendation that medications were largely ineffective for carpal tunnel syndrome.

Chiropractic Care for Carpal Tunnel Syndrome

Panelists agreed that the guidelines on chiropractic care for carpal tunnel syndrome were valid, with a median rating of eight. Two panelists found the recommendations to be appropriate because there is little evidence to support chiropractic care for carpal tunnel syndrome.

Nonoperative Procedures for Carpal Tunnel Syndrome

Panelists agreed that the guidelines on nonoperative procedures for carpal tunnel syndrome were valid, with a median rating of eight. Panelists found the recommendations for this section to be standard and well-referenced. For carpal tunnel syndrome, one panelist commented that

nonoperative procedures are primarily injections and that the one-injection recommendation is appropriate.

Surgery for Carpal Tunnel Syndrome

Panelists agreed that the guidelines on surgery for carpal tunnel syndrome were valid, with a median rating of eight. One panelist commented that the section on open and endoscopic approaches and surgeon experience was appropriately described. Another panelist found the summary of the indications and options for surgery to be well-structured.

Neck and Upper Back Injuries

Seven panelists provided ratings for the appropriateness of guidelines related to neck and upper back injuries, except for in the category of diagnostic imaging, in which only six panelists provided ratings. One category of management options was rated as uncertain (see Table 3.5).

Table 3.5. Ratings for Guidelines on Neck and Upper Back Injuries

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	7	Valid	8	4–8
Physical Therapy	7	Valid	8	4–8
Medication Therapy	6.5	Valid	8	5–8
Chiropractic Care	7	Valid	8	5–8
Nonoperative Procedures	5.5	Uncertain	8	3–7
Surgery	8	Valid	7	7–9

Diagnostic Imaging for Neck and Upper Back Injuries

Panelists agreed that the guidelines on diagnostic imaging for neck and upper back injuries were valid, with a median rating of seven. A few panelists found that the summaries appropriately discussed the range of radiological methods, including X-rays and MRIs. However, other panelists strongly suggested that the criteria for MRI were too stringent for workers with neurologic deficits. Suggestions included adding detail about timing and clinical features, including whiplash as a separate diagnostic category, and stating indications and contraindications for MRI more explicitly.

Physical Therapy for Neck and Upper Back Injuries

Panelists agreed that the guidelines on physical therapy for neck and upper back injuries were valid, with a median rating of seven. One panelist commented that the physical therapy summaries were appropriately categorised by diagnosis, while another panelist commented that the range of treatments was outlined and referenced. Panelists suggested including guidelines for splinting, spinal manipulation, and manual therapy (which includes more than just manual

traction). One panelist noted that the evidence basis for the number of sessions was unspecified and therefore unclear.

Medication Therapy for Neck and Upper Back Injuries

Panelists appraised the guidelines on medication therapy for neck and upper back injuries to be valid, with a median rating of 6.5. One panelist stated that the information presented was valid and relevant. Suggestions included consolidating steroid references into one entry and reconciling the conflicting recommendations for the duration of use for opioid treatment between this chapter and the pain chapter. One panelist commented that muscle relaxants should only be used for the short-term, acute stage. Additionally, panelists commented that there were too many references to other chapters in this chapter.

Chiropractic Care for Neck and Upper Back Injuries

Most panelists agreed that the guidelines on chiropractic care for neck and upper back injuries were valid, with a median rating of seven. One panelist commented that the summaries were detailed and included relevant studies. Another panelist thought that ODG's recommendation of a limited role for chiropractic care was appropriate. However, several panelists noted criticisms in other sections that are also applicable to this section: manipulation is inappropriately equated with chiropractic care, and chiropractic care may differ considerably in Australia compared with U.S. practices.

Nonoperative Procedures for Neck and Upper Back Injuries

Panelists were uncertain about the guidelines on nonoperative procedures for neck and upper back injuries, with a median rating of 5.5. There was disagreement regarding the use of epidural steroid injections. Several panelists believed that ODG appropriately recommended against this practice due to limited evidence and some suggestion of potential harm.

Surgery for Neck and Upper Back Injuries

Panelists agreed that the guidelines on surgery for neck and upper back injuries were valid, with a median rating of eight. Panelists generally found the guidelines and comparisons of various surgical methods to be appropriate and acceptable, such as the summary of disc replacement, plating, and anterior cervical discectomy and fusion. However, one panelist questioned why laminoplasty was categorised with discectomy and laminectomy.

Forearm, Wrist, and Hand Injuries

Eight panelists provided ratings for the appropriateness of guidelines related to forearm, wrist, and hand injuries. All categories of management options were rated as valid (see Table 3.6).

Table 3.6. Ratings for Guidelines on Forearm, Wrist, and Hand Injuries

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	7	Valid	8	6–8
Physical Therapy	7	Valid	8	5–7
Medication Therapy	7	Valid	8	5–8
Chiropractic Care	7.5	Valid	8	4–8
Nonoperative Procedures	7	Valid	8	5–9
Surgery	7	Valid	8	7–9

Diagnostic Imaging for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on diagnostic imaging for forearm, wrist, and hand injuries were valid, with a median rating of seven. Panelists observed that the guidelines for this section were reasonable, well-documented, and appropriate, especially considering the broad range of injuries and limited evidence; for example, one panelist commented that the recommendations for X-ray, MRI with or without arthrography, CT, and ultrasound were appropriate. However, another panelist commented that there may need to be additional consideration of carpal tunnel syndrome as a differential diagnosis.

Physical Therapy for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on physical therapy for forearm, wrist, and hand injuries were valid, with a median rating of seven. Generally, panelists considered the guidelines to be valid and appropriate, especially given the difficulty of summarising the range of conditions. However, concerns were mentioned about the reference to the low back chapter in the entry for physical therapy. One panelist pointed out that the guidelines might overlook other issues that can affect these injuries, such as an injury proximal to the injury of interest. Additionally, panelists noted that the number of recommended sessions lacked supporting evidence.

Medication Therapy for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on medication therapy for forearm, wrist, and hand injuries were valid with a median rating of seven. Overall, panelists observed that this section was appropriate, including the references to the pain chapter; for example, one panelist described the summaries of injections for conditions such as trigger fingers as appropriate. However, panelists suggested that the recommendations on glucosamine/chondroitin and opioids for severe pain be revised because they do not reflect varied results in literature.

Chiropractic Care for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on chiropractic care for forearm, wrist, and hand injuries were valid, with a median rating of 7.5. Panelists explained that the guidelines appropriately recommended against manipulative treatment.

Nonoperative Procedures for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on nonoperative procedures for forearm, wrist, and hand injuries were valid, with a median rating of seven. Panelists considered the guidelines generally acceptable, with no obvious omissions. One panelist pointed out that activity and work modifications are key components of treating these injuries, but are not described with respect to specific injuries. This panelist also criticised the recommendation against rest, suggesting that relative rest or avoidance may be helpful for certain conditions, such as soft tissue injuries.

Surgery for Forearm, Wrist, and Hand Injuries

Panelists agreed that the guidelines on surgery for forearm, wrist, and hand injuries were valid, with a median rating of seven. Panelists considered the recommendations in the section to be appropriate.

Hip and Pelvis Injuries

All panelists rated the validity of ODG's chapter on hip and pelvis injuries. Ratings varied meaningfully across topics; the section on nonoperative procedures received the lowest median ratings (see Table 3.7).

Table 3.7. Ratings for Guidelines on Hip Injuries

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	8	Valid	8	6–9
Physical Therapy	7	Valid	8	3–8
Medication Therapy	7	Valid	8	5–8
Chiropractic Care	6	Uncertain	8	6–7
Nonoperative Procedures	5.5	Uncertain	8	3–7
Surgery	7	Valid	8	7–8

Diagnostic Imaging for Hip and Pelvis Injuries

Panelists agreed that the guidelines on diagnostic imaging for hip and pelvis injuries were valid, with a median rating of eight. Overall, panelists thought that the recommendations were appropriate. Panelists commented that the summaries of X-rays, CT scans, and MRIs were well-documented and valid. Additionally, one panelist praised the recommendations for imaging for trauma and function.

Physical Therapy for Hip and Pelvis Injuries

Panelists agreed that the guidelines on physical therapy for hip and pelvis injuries were valid, with a median rating of seven. One panelist observed that the segmented analysis of physical therapy was especially appropriate. Similar to other sections, one panelist pointed out that there

is little evidence supporting rules that specify the number of recommended physical therapy sessions.

Medication Therapy for Hip and Pelvis Injuries

Panelists agreed that the guidelines on medication therapy for hip and pelvis injuries were valid, with a median rating of seven. Recommendations for prophylaxis and acetaminophen for first-line medication were considered generally valid. One panelist suggested clarifying the language on the justifications for using medications, such as using bisphosphonates regardless of bone density in hip fractures. Another panelist noted that the recommendation that opioids be considered for severe acute pain may not be consistent with other recent prescribing guidelines.

Chiropractic Care for Hip and Pelvis Injuries

Panelists were uncertain about the guidelines on chiropractic care for hip and pelvis injuries; they were given a median rating of six. The limited role of manipulation for hip and pelvis injuries was considered appropriate. However, as mentioned by a panelist in other chapters, ODG recommendations favour chiropractic care more strongly than would be typical in Australian practice, probably because ODG reflects U.S. clinical practice. This panelist also noted that recommendations regarding duration and frequency of treatment appear reasonable. One panelist suggested that the summary addressing manipulation in patients undergoing hip arthroplasty should be edited to reflect the referenced studies, which primarily refer to osteopathic manipulative therapy. Additional comments noted that manipulation is effective for those with sacroiliac pain and that the adverse effects of manipulation with anaesthesia should be described.

Nonoperative Procedures for Hip and Pelvis Injuries

Panelists were uncertain about the guidelines on nonoperative procedures for hip and pelvis injuries, with a median rating of 5.5. Panelists often commented that the guidelines were acceptable and valid, with one panelist appreciating the summary of bursal injections. However, panelists had several concerns. One panelist commented that the guidelines should clarify that platelet-rich plasma treatments are not indicated for hip problems at this time. Panelists expressed confusion as to why sacroiliac joint fusion would be recommended, but injections and radiofrequency ablation would not be recommended. There was also disagreement on the recommendations for piriformis injections. One panelist agreed with the ODG recommendation to allow piriformis injections, but another panelist remarked that piriformis syndrome was a controversial diagnosis with no accepted objective diagnostic methods, which would make diagnosis in a population of workers particularly difficult.

Surgery for Hip and Pelvis Injuries

Panelists agreed that the guidelines on surgery for hip and pelvis injuries were valid, with a median rating of seven. Generally, panelists found the guidelines to be valid and appropriately

documented, with some of the summaries including recent references. One panelist suggested that the recommendations for hip arthroscopy should note that evidence is lacking for most of the indications listed in the recommendations.

Chronic Pain

All panelists rated the validity of ODG's chapter on chronic pain. Ratings varied meaningfully across topics (see Table 3.8).

Table 3.8. Ratings for Guidelines on Chronic Pain

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	6.5	Valid	8	5–8
Physical Therapy	7	Valid	7	6–8
Medication Therapy	8	Valid	7	7–9
Chiropractic Care	6	Uncertain	8	5–8
Nonoperative Procedures	7	Valid	8	4–8
Surgery	5.5	Uncertain	8	1–8

Diagnostic Imaging for Chronic Pain

Panelists agreed that the guidelines on diagnostic imaging for chronic pain were valid, with a median rating of 6.5. Although the preface to this chapter defines chronic pain, panelists expressed concern that a chapter on chronic pain would be broad and overlap with other chapters, as chronic pain can be associated with a wide range of conditions.

Physical Therapy for Chronic Pain

Panelists agreed that the guidelines on physical therapy for chronic pain were valid, with a median rating of seven. Panelists generally found the guidelines to be appropriate, with one panelist labelling the guidelines for this section as "very comprehensive." In particular, panelists appreciated the recommendation of physical therapy followed by home exercise programs and the recommendations against passive treatments. However, panelists were concerned that ODG recommended specific numbers of sessions without evidence. In addition, one panelist gave this section a lower rating due to the poor quality of evidence supporting the recommendation for whole-body vibration exercise, which included incomplete studies and studies with small sample sizes.

Medication Therapy for Chronic Pain

Panelists agreed that the guidelines on medication therapy for chronic pain were valid, with a median rating of eight. Several panelists mentioned the summaries covering opioids and generally thought that the guidelines covered a broad topic fairly well, including advising against

opioids as a first-line treatment. However, several panelists expressed concern that the opioid summaries were open to interpretation and could be used to support differing perspectives, perhaps due to a lack of emphasis on alternative treatment options.

Chiropractic Care for Chronic Pain

Panelists were uncertain about the guidelines on chiropractic care for chronic pain, with a median rating of six. Some panelists noted that guidelines were well-supported and in line with accepted practice, including the eight-session limit. However, another panelist pointed out that the summary for chiropractic care in chronic pain covered physical medicine treatments, not just manipulation. This approach was noted to be inconsistent with other chapters that equated chiropractic care with diverse forms of manipulation; the panelist suggested that the approach to chiropractic care should be consistent across chapters.

Nonoperative Procedures for Chronic Pain

Panelists agreed that the guidelines on nonoperative procedures for chronic pain were valid, with a median rating of seven. Overall, panelists generally commented that the guidelines were adequate, valid, and appropriate. One panelist recommended including walking programs.

Surgery for Chronic Pain

Panelists were uncertain about the guidelines on surgery for chronic pain, with a median rating of 5.5. Panelists explained that this section of the guidelines was limited; references were provided to other chapters. Panelists commented that surgery seemed primarily limited to spinal cord stimulation. One panelist found the spinal cord stimulation section to be well-researched; however, another panelist expressed concerns about the potential for adverse events from spinal cord stimulation and the lack of evidence presented for spinal cord stimulation for working non-cancer patients.

Mental Illness and Stress

Interventions for mental illness and stress received eight ratings per category; in general, these guidelines received lower ratings than those in most other sections (see Table 3.9). One panelist noted the section's impressive breadth but noted several important issues. First, ODG repeatedly references the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV, but most mental health practitioners use the updated version (DSM-V). Second, ODG arranges its recommendations by intervention without specifically mentioning particular clinical conditions (e.g., major depression). The panelist stated that this organisational structure reduces clinical validity because some treatments may work for some mental health conditions but not for others. Third, ODG uses overly broad and imprecise language at times: "psychological assessments" and "psychological testing" are used interchangeably, and "stress" is a symptom, but is treated by ODG as a valid diagnosis in itself. Fourth, the panelist believed that ODG implied that mental

health conditions lacked evidence on causation/aetiology; the panelist noted that at least some types of mental illness, such as posttraumatic stress syndrome (PTSD), have triggering events and/or biological pathways. Finally, the panelist noted that ODG conflated the difficulty of determining work-relatedness of mental illnesses with the difficulty of treating mental disorders. According to the panelist, although it may be difficult to determine whether some mental health disorders are work-related, this section uses the issue to undermine the efficacy of treatments for particular mental disorders, but does not cite sufficient supporting evidence.

Table 3.9. Ratings for Guidelines on Mental Illness and Stress

Category	Median	Decision	Number of Ratings	Range
Diagnostic Imaging	6	Uncertain	8	3–9
Psychological Therapies	7	Valid	8	3–7
Medication Therapy	7	Valid	8	3–8
Nonoperative Procedures	6.5	Valid	8	4–8

Diagnostic Imaging for Mental Illness and Stress

Panelists were uncertain about the guidelines on diagnostic imaging for mental illness and stress, with a median rating of six. Panelists appreciated the breadth of the topic, and one agreed with the emphasis on accurate diagnosis and recognition of the challenges. The panelists noted that ODG recommended appropriate tests, such as the Minnesota Multiphasic Personality Inventory and Brief Battery for Health Improvement 2. One panelist commented that recommendations for most of the psychological tests relied on literature examining their use in assessing chronic pain, rather than broader mental health conditions. Additionally, ODG included assessments for pain, and one panelist thought that this topic was outside the scope of a chapter on mental illness and stress.

Psychological Therapies for Mental Illness and Stress

Panelists appraised the guidelines on psychological therapies for mental illness and stress to be valid, with a median rating of seven. However, panelists thought the quality of the recommendations varied significantly by topic. Panelists commented that the summaries seemed to appropriately describe and reference the relevant therapies, such as cognitive behavioural therapy. However, one panelist commented that the recommended limit on the quantity of cognitive behavioural therapy sessions did not seem to be based on the evidence. Another panelist noted that phone-based therapy and bibliotherapy had less evidence than the guidelines acknowledged.

Medication Therapy for Mental Illness and Stress

Panelists appraised the guidelines on medication therapy for mental illness and stress as valid, with a median rating of seven. Several panelists commented that the guidelines seemed to be well-referenced. However, concerns were expressed about the clarity of the evidence base and variability in the quality of the recommendations. Additionally, some sections were noted to be out of date. For example, the summary of PTSD pharmacotherapy did not acknowledge the expert consensus that atypical antipsychotics have a useful role as augmenting agents, especially for treatment-refractory PTSD. In addition, recommendations are grouped together inappropriately; for example, ODG states that atypical antipsychotic drugs are not recommended, but does not note for which indication the treatment is ineffective.

Nonoperative Procedures for Mental Illness and Stress

Panelists appraised the guidelines on nonoperative procedures for mental illness and stress as valid, with a median rating of 6.5. Panelists agreed with the recommendations, especially for electroconvulsive therapy. Another suggested including health coaching as a potential treatment option. One panelist pointed out that the summaries equate acupuncture with counselling, when acupuncture is generally an adjunctive treatment for moderate-to-severe depression.

Fitness for Duty

Panelists agreed that the guidelines on fitness for duty (i.e., return to work) were valid, with a median rating of seven. Most panelists noted that the recommendations seemed to be thoughtful, with one panelist commenting that the section was easy to read; another observed that references to other chapters were appropriate. In addition, several panelists commented that the inclusion of different occupations in the summaries seemed haphazard, with significant variation in the level of detail and quality of the evidence. Another panelist suggested that ODG include vocational rehabilitation, physical therapy, and cognitive-behavioural programs.

In addition, one panelist suggested that the selection of professions and the relationship to jurisdictional requirements was more relevant for a U.S. audience, and recommended including an Australian study on fitness to drive. Another mentioned that providers need to follow jurisdictional requirements (which, per the panelist, ODG did not mention).

Differences by Specialty

Panelists rated a wide variety of topics, some of which were closer and some of which were further outside of their direct field of expertise. To determine whether differences in specialty may have affected the results, we show sensitivity analyses for topics appraised as uncertain by the entire panel. Table 3.10 compares results from the full panel with those derived from panelists with greater expertise on a given topic. Although the smaller sample sizes (two to four panelists) and lack of statistical testing does not allow us to draw definitive conclusions,

panelists having somewhat greater expertise did not appear to provide more or less favourable ratings for these six clinical areas than the full panel provided.

Table 3.10. Comparison of Full Panel vs. Expert-only Assessments for Topics Rated Uncertain

Topic	Category	Median (Full Panel)	Median (Panelists with Greater Expertise)
Hip and Pelvis Injuries	Chiropractic Therapies	6	7
Hip and Pelvis Injuries	Surgery	5.5	6
Chronic Pain	Chiropractic Therapies	6	6.5
Chronic Pain	Surgery	5.5	5
Mental Illness and Stress	Diagnostic Procedures	6	4.5
Neck and Upper Back Injuries	Nonoperative Procedures	5.5	4

Panelists with somewhat greater expertise in topics related to mental health provided qualitatively lower ratings on the mental illness and stress chapter. However, the panel included few individuals with expertise in psychiatry; as a result, we are uncertain as to how much the results would have changed with a different composition of panelists. Furthermore, we were unable to include a chiropractor on the panel and cannot be sure how individuals who practice chiropractic care would view the guideline content.

4. Conclusion

Main Findings

Payers in workers' compensation systems frequently use the ODG part of their utilisation review procedures. The objective of the current work was to evaluate the quality of the ODG, including technical quality, meaning the rigour of development methods, as well as clinical acceptability, meaning the validity of the guideline in the eyes of diverse clinical experts.

To assess technical quality, a team of four trained appraisers applied the AGREE II and AMSTAR instruments to evaluate guideline development methods and supporting systematic reviews, respectively. The overall score on the AGREE II was 58 per cent (scale: 0–100 per cent), and the overall score on the AMSTAR was "fair/good" (scale: poor to outstanding). The ODG's strengths included an expansive scope, clearly written recommendations, frequent updating, and the associated development of a well-designed tool for applying recommendations to clinical practice. Important weaknesses included lack of input from workers with occupational conditions and limited information about the current process by which ODG chapter development teams identify, select, evaluate, and synthesise evidence. In addition, appraisers were uncertain whether experienced methodologists were involved in development and whether ODG chapter development teams were free of conflicts of interest and had editorial independence from WLDI.

Our AGREE II and AMSTAR appraisals reveal several specific ways in which the technical quality of ODG could be improved:

1. engaging representatives of workers/labour in guideline development
2. involving experienced methodologists in all aspects of guideline development, particularly in conducting literature searches and synthesising evidence
3. improving documentation of search terms, selection criteria, and numbers of studies identified by the searches and eligible for evidence syntheses
4. creating evidence tables that describe published literature for each recommendation
5. documenting and mitigating any conflicts of interest among ODG chapter development teams and ensuring that they have editorial independence.

To assess ODG's clinical acceptability, a multidisciplinary panel of expert clinicians from the United States and Australia rated the validity of selected guideline content in nine ODG chapters on common occupational conditions. Panelists were instructed that validity meant consistency with published evidence and expert opinion as well as the greatest likelihood of offering patients a favourable balance of risks and benefits. Overall, the panelists found selected content in most of these chapters to be valid (41 of 47 topics), although some content was judged to be of uncertain validity (six of 47 topics). For important topics (low back pain; shoulder injuries; carpal tunnel syndrome; forearm, wrist, and hand injuries; and fitness for duty),

panelists rated all relevant content to be valid. For the remaining topics, most of the selected content was judged valid. For three important topics, including medications and nonoperative procedures for neck and upper back injuries, as well as nonoperative procedures for hip injuries, content was of uncertain validity. Several panelists also noted that the evidence basis for numbers of allowed physical therapy sessions was uncertain; the recommendations are based on U.S. practice and they may have less utility in non-U.S. settings. Content was also of uncertain validity for five topics where the scope of the relevant material was limited, including chiropractic care for hip injuries and chronic pain, surgery for chronic pain, and diagnostic imaging and psychological therapies for mental illness and stress. Uncertainty related to the validity of content on mental illness and chiropractic care may have been affected by the fact that the panel included only one mental health professional and did not include a chiropractor. Panelists made comments and suggestions for improvement, many of which identified content that extended beyond the published evidence or that warranted updating.

Limitations

This evaluation has several limitations. First, we did not independently review the primary literature on the many topics covered by ODG, so we cannot say with certainty that the guideline is evidence-based. Our evaluation instead focuses on methods of development and acceptability to clinical experts. Second, the ODG has a broad scope, covering most occupational conditions, as well as treatments that range from widely accepted to rarely used and highly controversial. Panelists were unable to rate every individual recommendation included in the ODG because of the broad scope. Instead, they rated the validity of broad aspects of care for specific conditions and had to make judgements about how to prioritise the material. This also made it harder to ensure that the panel composition matched the full spectrum of care covered by the guidelines. Nonetheless, the panel included individuals with substantial expertise related to musculoskeletal conditions and chronic pain, which are common among workers with occupational disorders. Third, the panel included more experts from the United States than from Australia, and the guideline was designed for use in the United States. However, ratings were similar between United States and Australian panelists (aside from slight differences in ratings for chiropractic care, which could be affected by differences in the practice of chiropractic care between the countries).

Implications

It may be helpful to put these findings in the context of prior evaluations as well as recommended standards for developing guidelines and performing systematic reviews. Using similar methods for evaluating technical quality, we recently assessed 13 guidelines on the use of opioids for chronic pain, including the ODG and another workers' compensation guideline. In this prior work, ODG received an AGREE II score of 51 per cent and an AMSTAR score of

"fair/good" (Nuckols et al., 2014). Therefore, the current and prior ODG scores are similar, despite the fact that the prior scores were based on publicly available information (which was limited for ODG), while present scores were based on detailed information provided by WLDI.

More importantly, ODG's technical quality in that study was on par with the other opioid guidelines examined, although ODG fell short of the highest quality opioid guidelines evaluated. Across the 13 guidelines, overall AGREE II scores had a mean of 52 per cent and a range from 28 per cent to 76 per cent, while overall AMSTAR scores had a median of "fair" and a range from "poor" to "excellent/outstanding." Scores on the six domains of the AGREE II were also on par with those of the other guidelines, including scope and purpose (64 per cent for ODG vs. group mean of 69 per cent and range 39–89 per cent), rigour of development (52 per cent vs. group mean 48 per cent and range 20–84 per cent), clarity of presentation (75 per cent vs. group mean of 71 per cent and range 37–93 per cent), editorial independence (69 per cent vs. group mean of 44 per cent, and range 0–88 per cent), stakeholder involvement (67 per cent vs. group mean of 52 per cent and range 23–77 per cent), and applicability (74 per cent vs. group mean of 37 per cent and range 13–56 per cent) (Nuckols et al., 2014).

As noted previously, the Institute of Medicine has issued recommended standards for developing guidelines and performing systematic reviews. While the AGREE II is generally well aligned with the current Institute of Medicine standards for guideline development, the AMSTAR is not as strict as the Institute of Medicine current standards for performing systematic reviews, which were published more recently than the AMSTAR (Shea et al., 2009; Brouwers et al., 2010a; Institute of Medicine, 2011b). Adhering to standards by the Institute of Medicine would enable WLDI to further increase the technical quality of the ODG. Improvements in technical quality should increase the likelihood that clinicians find guideline recommendations to be credible; adhering to those recommendations should, in turn, improve patient outcomes.

Our current evaluation of clinical acceptability differs somewhat from our prior evaluation of five guideline sets for occupational disorders in 2004. First, we used different methods: current panelists rated more topics (47 instead of ten); participated in an online, not in-person, rating process; and focused on one guideline, rather than five guidelines. Second, current assessments of clinical acceptability appear much more favourable than prior ones; in the last study, the panelists found content to be valid for only four of ten topics (Nuckols et al., 2005). Methodological differences seem unlikely to fully account for the more favourable findings. We suspect that ODG may have received higher scores in the present evaluation because the guideline regularly accepts and solicits input from clinicians.

For policymakers and payers considering whether to adopt ODG, other considerations may be relevant in addition to technical quality and clinical acceptability. In prior publications, we have discussed at length the considerations involved in choosing and implementing guidelines in workers' compensation settings (Nuckols et al., 2005; Harber et al., 2008). Today, few, if any, other occupational medicine guidelines have been specifically designed for use in utilisation review. For example, the American College of Occupational and Environmental Medicine

guideline also addresses a wide range of occupational disorders, but it was designed to support clinical decisionmaking, although developers do acknowledge it is sometimes applied for utilisation review (American College of Environmental and Occupational Medicine, 2016). In contrast, ODG specifically includes tools that payers can use to look up the appropriateness of proposed treatments, with links to evidence summaries. To apply ODG in Australia, some adaptation may be helpful, particularly with regard to names of medications, use of chiropractic care, and possibly limits on the number of physical therapy sessions. The alternative to using an existing guideline is developing one, but this would take several years and be prohibitively expensive due to the broad scope of occupational conditions.

In conclusion, our analysis found that the technical quality of the ODG has some limitations, but that clinical experts in diverse fields from two countries considered content related to several important occupational conditions to be valid. As WLDI updates ODG, methodological improvements are warranted to ensure that the guideline continues to keep pace with advances in the science of guideline development as well as the expanding base of literature. In this way, applying the ODG in utilisation review procedures is likely to yield the best possible clinical outcomes in addition to reducing unnecessary healthcare expenditures.

Appendix

National Academy Recommended Standards for Initiating a Systematic Review

Standard 2.1: Establish a team with appropriate expertise and experience to conduct the systematic review[a]	2.1.1. Include expertise in the pertinent clinical content areas 2.1.2. Include expertise in systematic review methods 2.1.3. Include expertise in searching for relevant evidence 2.1.4. Include expertise in quantitative methods 2.1.5. Include other expertise as appropriate
Standard 2.2: Manage bias and COIs of the team conducting the systematic review	2.2.1. Require each team member to disclose potential COIs and professional or intellectual bias 2.2.2. Exclude individuals with a clear financial conflict 2.2.3. Exclude individuals whose professional or intellectual bias would diminish the credibility of the review in the eyes of the intended users
Standard 2.3: Ensure user and stakeholder input as the review is designed and conducted	2.3.1. Protect the independence of the review team to make the final decisions about the design, analysis, and reporting of the review
Standard 2.4: Manage bias and COIs for individuals providing input into the systematic review	2.4.1. Require individuals to disclose potential COIs and professional or intellectual bias 2.4.2. Exclude input from individuals whose COIs or bias would diminish the credibility of the review in the eyes of the intended users
Standard 2.5: Formulate the topic for the systematic review	2.5.1. Confirm the need for a new review 2.5.2. Develop an analytic framework that clearly lays out the chain of logic that links the health intervention to the outcomes of interest and defines the key clinical questions to be addressed by the systematic review 2.5.3. Use a standard format to articulate each clinical question of interest 2.5.4. State the rationale for each clinical question 2.5.5. Refine each question based on user and stakeholder input
Standard 2.6: Develop a systematic review protocol	2.6.1. Describe the context and rationale for the review from both a decisionmaking and research perspective 2.6.2. Describe the study screening and selection criteria (inclusion/exclusion criteria) 2.6.3. Describe precisely which outcome measures, time points, interventions, and comparison groups will be addressed 2.6.4. Describe the search strategy for identifying relevant evidence 2.6.5. Describe the procedures for study selection 2.6.6. Describe the data extraction strategy 2.6.7. Describe the process for identifying and resolving disagreement between researchers in study selection and data extraction decisions 2.6.8. Describe the approach to critically appraising individual studies 2.6.9. Describe the method for evaluating the body of evidence, including the quantitative and qualitative synthesis strategies 2.6.10. Describe and justify any planned analyses of differential treatment effects according to patient subgroups, how an intervention is delivered, or how an outcome is measured 2.6.11. Describe the proposed timetable for conducting the review
Standard 2.7: Submit the protocol for peer review	2.7.1. Provide a public comment period for the protocol and publicly report on disposition of comments
Standard 2.8: Make the final protocol publicly available, and add any amendments to the protocol in a timely fashion	

Standards for Finding and Assessing Individual Studies	
Standard 3.1: Conduct a comprehensive systematic search for evidence	3.1.1. Work with a librarian or other information specialist trained in performing systematic reviews to plan the search strategy 3.1.2. Design the search strategy to address each key research question 3.1.3. Use an independent librarian or other information specialist to peer review the search strategy 3.1.4. Search bibliographic databases 3.1.5. Search citation indexes 3.1.6. Search literature cited by eligible studies 3.1.7. Update the search at intervals appropriate to the pace of generation of new information for the research question being addressed 3.1.8. Search subject-specific databases if other databases are unlikely to provide all relevant evidence 3.1.9. Search regional bibliographic databases if other databases are unlikely to provide all relevant evidence
Standard 3.2: Take action to address potentially biased reporting of research results	3.2.1. Search grey literature databases, clinical trial registries, and other sources of unpublished information about studies 3.2.2. Invite researchers to clarify information about study eligibility, study characteristics, and risk of bias 3.2.3. Invite all study sponsors and researchers to submit unpublished data, including unreported outcomes, for possible inclusion in the systematic review 3.2.4. Hand search selected journals and conference abstracts 3.2.5. Conduct a web search 3.2.6. Search for studies reported in languages other than English if appropriate
Standard 3.3: Screen and select studies	3.3.1. Include or exclude studies based on the protocol's prespecified criteria 3.3.2. Use observational studies in addition to randomised clinical trials to evaluate harms of interventions 3.3.3. Use two or more members of the review team, working independently, to screen and select studies 3.3.4. Train screeners using written documentation; test and retest screeners to improve accuracy and consistency 3.3.5. Use one of two strategies to select studies: (1) read all full-text articles identified in the search or (2) screen titles and abstracts of all articles and then read the full text of articles identified in initial screening 3.3.6. Taking account of the risk of bias, consider using observational studies to address gaps in the evidence from randomised clinical trials on the benefits of interventions
Standard 3.4: Document the search	3.4.1. Provide a line-by-line description of the search strategy, including the date of every search for each database, web browser, etc. 3.4.2. Document the disposition of each report identified including reasons for their exclusion if appropriate
Standard 3.5: Manage data collection	3.5.1. At a minimum, use two or more researchers, working independently, to extract quantitative and other critical data from each study. For other types of data, one individual could extract the data while the second individual independently checks for accuracy and completeness. Establish a fair procedure for resolving discrepancies—do not simply give final decisionmaking power to the senior reviewer 3.5.2. Link publications from the same study to avoid including data from the same study more than once 3.5.3. Use standard data extraction forms developed for the specific systematic review 3.5.4. Pilot-test the data extraction forms and process
Standard 3.6: Critically appraise each study	3.6.1. Systematically assess the risk of bias, using predefined criteria 3.6.2. Assess the relevance of the study's populations, interventions, and outcome measures 3.6.3. Assess the fidelity of the implementation of interventions

Standards for Synthesizing the Body of Evidence	
Standard 4.1: Use a prespecified method to evaluate the body of evidence	4.1.1. For each outcome, systematically assess the following characteristics of the body of evidence: risk of bias, consistency, precision, directness, reporting bias 4.1.2. For bodies of evidence that include observational research, also systematically assess the following characteristics for each outcome: dose-response association, plausible confounding that would change the observed effect, strength of association 4.1.3. For each outcome specified in the protocol, use consistent language to characterise the level of confidence in the estimates of the effect of an intervention
Standard 4.2: Conduct a qualitative synthesis	4.2.1. Describe the clinical and methodological characteristics of the included studies, including their size, inclusion or exclusion of important subgroups, timeliness, and other relevant factors 4.2.2. Describe the strengths and limitations of individual studies and patterns across studies 4.2.3. Describe, in plain terms, how flaws in the design or execution of the study (or groups of studies) could bias the results, explaining the reasoning behind these judgements 4.2.4. Describe the relationships between the characteristics of the individual studies and their reported findings and patterns across studies 4.2.5. Discuss the relevance of individual studies to the populations, comparisons, co-interventions, settings, and outcomes or measures of interest
Standard 4.3: Decide if, in addition to a qualitative analysis, the systematic review will include a quantitative analysis (meta-analysis)	4.3.1. Explain why a pooled estimate might be useful to decisionmakers
Standard 4.4: If conducting a meta-analysis, then do the following:	4.4.1. Use expert methodologists to develop, execute, and peer review the meta-analyses 4.4.2. Address the heterogeneity among study effects 4.4.3. Accompany all estimates with measures of statistical uncertainty 4.4.4. Assess the sensitivity of conclusions to changes in the protocol, assumptions, and study selection (sensitivity analysis)
Standards for Reporting Systematic Reviews	
Standard 5.1: Prepare final report using a structured format	5.1.1. Include a report title 5.1.2. Include an abstract 5.1.3. Include an executive summary 5.1.4. Include a summary written for the lay public 5.1.5. Include an introduction (rationale and objectives) 5.1.6. Include a methods section. Describe the following: • Research protocol • Eligibility criteria (criteria for including and excluding studies in the systematic review) • Analytic framework and key questions • Databases and other information sources used to identify relevant studies • Search strategy • Study selection process • Data extraction process • Methods for handling missing information • Information to be extracted from included studies • Methods to appraise the quality of individual studies • Summary measures of effect size (e.g., risk ratio, difference in means) • Rationale for pooling (or not pooling) results of included studies • Methods of synthesising the evidence (qualitative and meta-analysis) • Additional analyses, if done, indicating which were prespecified

	5.1.7. Include a results section. Organise the presentation of results around key questions. Describe the following (repeat for each key question): • Study selection process • List of excluded studies and reasons for their exclusion • Appraisal of individual studies' quality • Qualitative synthesis • Meta-analysis of results, if performed (explain rationale for doing one) • Additional analyses, if done, indicating which were prespecified • Tables and figures 5.1.8. Include a discussion section. Include the following: • Summary of the evidence • Strengths and limitations of the systematic review • Conclusions for each key questions • Gaps in evidence • Future research needs 5.1.9. Include a section describing funding sources and COIs
Standard 5.2: Peer review the draft report	5.2.1. Use a third party to manage the peer review process 5.2.2. Provide a public comment period for the report and publicly report on disposition of comments
Standard 5.3: Publish the final report in a manner that ensures free public access	

[a] As per the Institute of Medicine document, numbering starts at 2.1.

References

American College of Occupational and Environmental Medicine, "The Role of ACOEM's Practice Guidelines in Treatment Decisions," web page, undated. As of December 12, 2016: http://www.acoem.org/responsetoRSDSA.aspx

Brouwers, M. C., M. E. Kho, G. P. Browman, J. S. Burgers, F. Cluzeau, G. Feder, B. Fervers, I. D. Graham, J. Grimshaw, S. E. Hanna, P. Littlejohns, J. Makarski, L. Zitzelsberger, and AGREE Next Steps Consortium, "AGREE II: Advancing Guideline Development, Reporting and Evaluation in Health Care," *Canadian Medical Association Journal*, Vol. 182, No. 18, December 14, 2010a, pp. E839–842.

———, "Development of the AGREE II, Part 1: Performance, Usefulness and Areas for Improvement," *Canadian Medical Association Journal*, Vol. 182, No. 10, July 13, 2010b, pp. 1045–1052.

———, "Development of the AGREE II, Part 2: Assessment of Validity of Items and Tools to Support Application," *Canadian Medical Association Journal*, Vol. 182, No. 10, July 13, 2010c, pp. E472–478.

Colorado Department of Labor and Employment, "Utilization Standards/Medical Billing Disputes," web page, undated. As of December 12, 2016: https://www.colorado.gov/pacific/cdle/utilization-standardsmedical-billing-disputes

Fitch, Kathryn, Steven J. Bernstein, Maria Dolores Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam von het Loo, Joseph McDonnell, Janneke Vader, and James P. Kahan, *The RAND/UCLA Appropriateness Method User's Manual*, Santa Monica, Calif.: RAND Corporation, MR-1269-DG-XII/RE, 2001 As of December 29, 2016: http://www.rand.org/pubs/monograph_reports/MR1269.html

Harber, Philip, Barbara O. Wynn, Yee-Wei Lim, Soeren Mattke, Steven M. Asch and Teryl K. Nuckols, "Selection of Workers' Compensation Treatment Guidelines: California Experience," *Journal of Occupational and Environmental Medicine*, Vol. 50, No. 11, November 2008, pp. 1282–1292.

Hunt, D. L., J. Tower, R. D. Artuso, J. A. White, C. Bilinski, J. Rademacher, X. Tao, and E. J. Bernacki, "A New Method of Assessing the Impact of Evidence-Based Medicine on Claim Outcomes," *Journal of Occupational and Environmental Medicine*, Vol. 58, No. 5, May 2016, pp. 519–524.

Institute of Medicine of the National Academies, *Controlling Costs and Changing Patient Care? The Role of Utilization Management*, B. H. Gray and M. J. Field, eds., Washington, D.C.: National Academies Press, 1989.

———, *Clinical Practice Guidelines We Can Trust*, Robin Graham, Michelle Mancher, Dianne Miller Wolman, Sheldon Greenfield, and Earl Steinberg, eds., Washington D.C.: National Academies Press, 2011a.

———, *Finding What Works in Health Care: Standards for Systematic Reviews*, Jill Eden, Laura Levit, Alfred Berg, and Sally Morton, Washington, D.C.: National Academies Press, 2011b.

Khodyakov, D., L. Mikesell, R. Schraiber, M. Booth, and E. Bromley, "On Using Ethical Principles of Community-Engaged Research in Translational Science," *Translational Research*, Vol. 171, 2016, pp. 52–62.

Nuckols, Teryl K., Laura Anderson, Ioana Popescu, Allison L. Diamant, Brian Doyle, Pail Di Capua, and Roger Chou, "Opioid Prescribing: A Systematic Review and Critical Appraisal of Guidelines for Chronic Pain," *Annals of Internal Medicine*, Vol. 160, No. 1, January 7, 2014, pp. 38–47.

Nuckols, Teryl K., Yee-Wei Lim, Barbara O. Wynn, Soeren Mattke, Catherine H. MacLean, Philip Harber, Robert H. Brook, Peggy Wallace, Rena H. Garland and Steven Asch, "Rigorous Development Does Not Ensure That Guidelines Are Acceptable to a Panel of Knowledgeable Providers," *Journal of General Internal Medicine*, Vol. 23, No. 1, January 23, 2008, pp. 37–44.

Nuckols, Teryl K., Barbara O. Wynn, Yee-Wei Lim, Rebecca Shaw, Soeren Mattke, Thomas Wickizer, Phillip Harber, Peggy Wallace, Steven M. Asch, Catherine MacLean, and Rena Hasenfeld Garland, *Evaluating Medical Treatment Guideline Sets for Injured Workers in California*, Santa Monica, Calif: RAND Corporation, MG-400-ICJ, 2005. As of December 14, 2016: http://www.rand.org/pubs/monographs/MG400.html

Pennsylvania Department of Labor and Industry, "Health Care Services Review FAQs: Utilization Review," web page, undated. As of December 12, 2016: http://www.dli.pa.gov/Businesses/Compensation/WC/faq/Pages/Health-Care-Services-Review-FAQs.aspx

Sengupta, Ishita, and Marjorie L. Baldwin, *Workers' Compensation: Benefits, Coverage, and Costs, 2013*, Washington, D.C.: National Academy of Social Insurance, 2015.

Shea, B. J., C. Hamel, G. A. Wells, L. M. Bouter, E. Kristjansson, J. Grimshaw, D. A. Henry, and M. Boers, "AMSTAR Is a Reliable and Valid Measurement Tool to Assess the

Methodological Quality of Systematic Reviews," *Journal of Clinical Epidemiology*, Vol. 62, No. 10, October 2009, pp. 1013–1020.

Shekelle, P. G., E. Ortiz, S. Rhodes, S. C. Morton, M. P. Eccles, J. M. Grimshaw, and S. H. Woolf, "Validity of the Agency for Healthcare Research and Quality Clinical Practice Guidelines: How Quickly Do Guidelines Become Outdated?" *Journal of the American Medical Association*, Vol. 286, No. 12, September 26, 2001, pp. 1461–1467.

Shojania, K. G., M. Sampson, M. T. Ansari, J. Ji, S. Doucette, and D. Moher, "How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis," *Annals of Internal Medicine*, Vol. 147, No. 4, August 21, 2007, pp. 224–233.

State of California Department of Industrial Relations, "Utilization Review," web page, August 2016. As of December 12, 2016: http://www.dir.ca.gov/dwc/UR_Main.htm

Texas Department of Insurance, "Review of Medical Necessity (Utilization Review)," web page, September 6, 2014. As of December 12, 2016: http://www.tdi.texas.gov/wc/carrier/ur.html

Washington State Department of Labor and Industries, "Utilization Review," web page, undated. As of December 12, 2016: http://www.lni.wa.gov/ClaimsIns/Providers/AuthRef/UR.asp

Wickizer, T. M., G. Franklin, J. V. Gluck, and D. Fulton-Kehoe, "Improving Quality Through Identifying Inappropriate Care: The Use of Guideline-Based Utilization Review Protocols in the Washington State Workers' Compensation System" *Journal of Occupational and Environmental Medicine*, Vol. 46, No. 3, March 2004, pp. 198–204.

Wickizer, T. M., and D. Lessler, "Utilization Management: Issues, Effects, and Future Prospects," *Annual Review of Public Health*, Vol. 23, 2002, pp. 233–254.

Wickizer, T. M., D. Lessler, and G. Franklin, "Controlling Workers' Compensation Medical Care Use and Costs Through Utilization Management," *Journal of Occupational and Environmental Medicine*, Vol. 41, No. 8, August 1999, pp. 625–631.

Work Loss Data Institute, "Evidence-Based Medical Treatment and Return to Work Guidelines," web page, undated. As of August 30, 2016: http://www.worklossdata.com/