AJKD

Core Curriculum in Nephrology

Development of Clinical Practice Guidelines in the Care of People With Kidney Disease: Core Curriculum 2016

Angela C. Webster, PhD,[1,2,3] Evi V. Nagler, MD,[4,5] and Martin Gallagher, PhD[6,7]

DEFINITION, PURPOSE, AND EFFECTIVENESS OF CLINICAL PRACTICE GUIDELINES

Definition and Purpose

The Institute of Medicine defines clinical practice guidelines (hereafter referred to as "guidelines") as "statements that include recommendations intended to optimize patient care that are informed by a systematic of evidence and an assessment of the benefits and harms of alternative care options" (iom.nationalacademies.org/reports/2011/clinical-practice-guidelines-we-can-trust/report-brief.aspx).

The use of guidelines began and expanded in response to issues that all health care systems continue to face. Increasing amounts of information were gradually overwhelming clinicians who were finding it increasingly difficult to evaluate new information in context and provide the best possible care supported by the latest scientific developments. Health care costs were increasing continuously, service delivery varied among providers, and at least some of this variation reflected inappropriate care (both over- and underuse of services) with resultant effects on patient outcomes. Guidelines were developed in an attempt to remedy these problems by providing a quality control tool for standardizing and optimizing care.

Use, Expected Benefits, and Possible Harms

The ultimate aim of guidelines is to improve patient outcomes by optimizing medical decision making. Initially, guidelines were conceived with the primary goals of helping individual clinicians digest new research and facilitating evidence-based management decisions. By presenting contemporary evidence summaries and treatment recommendations, guideline entities aimed to improve effective practice by encouraging treatments with proven benefit and discouraging ineffective or harmful treatments.

Guidelines serve other purposes as well. They may have a general educational role, teaching patients, caregivers, and the media about best health care practices. Knowledge of a guideline may empower the individual patient to make better informed health care choices by promoting shared decision making between users and providers of health care. Guidelines generated for conditions with little or no research evidence may be useful in additional ways. Rather than emphasizing what we know, these guidelines may predominantly serve to highlight what we do not know, providing justification for new research plans and encouraging stake holders and funding agencies to support efforts that will close evidence gaps. Guidelines may also be used as an implementation tool to promote translation of research evidence into everyday clinical practice. Health care service providers may use guidelines as a tool for standardizing care, aiming to increase efficiency with a view to reducing costs. Guidelines may be used to derive measurable clinical performance indicators, which can be used in quality assurance initiatives, or may form the basis of financial reimbursement schemes by rewarding the use of best evidence-based practice. Publication of guidelines may raise public awareness of a previously underrecognized health burden and inform decisions by policy makers in allocating health care resources. Finally, clinicians may turn to guidelines for medicolegal protection in cases of malpractice litigation.

The potential benefits of guidelines rest on the premise that guidelines are valid and their recommendations are a credible distillation of available evidence. History has often challenged this assumption. Recommendations may simply be wrong if guideline panels do not appraise all the existing evidence, if they give undue weight to certain findings, or if the available evidence is misleading or misinterpreted. Recommendations within guidelines may be one sided if development groups fail to convene multistakeholder multidisciplinary panels or if panel

From the [1]Sydney School of Public Health, University of Sydney, Sydney; [2]Centre for Transplant and Renal Research, Westmead Hospital; [3]Cochrane Kidney and Transplant, The Children's Hospital at Westmead, Westmead, NSW, Australia; [4]Renal Division, Department of Internal Medicine, Ghent University Hospital, Ghent, Belgium; [5]European Renal Best Practice (ERBP), guidance body of the European Renal Association–European Dialysis and Transplant Association (ERA-EDTA), http://www.european-renal-best-practice.org/; [6]Concord Clinical School, University of Sydney; and [7]The George Institute for Global Health, Sydney, NSW, Australia.

Received April 9, 2015. Accepted in revised form August 8, 2015. Originally published online November 23, 2015.

Address correspondence to Angela C. Webster, PhD, Sydney School of Public Health, Edward Ford Building A27, University of Sydney, NSW 2006, Australia. E-mail: angela.webster@sydney.edu.au

0272-6386

http://dx.doi.org/10.1053/j.ajkd.2015.08.030

members are biased through financial or intellectual conflicts of interest.

Valid recommendations are never applicable to every patient, but it is a common misconception that guideline developers intend them to be so. If guideline recommendations are applied injudiciously, the frequently advertised benefit of more consistent practice (and thus reduced variation) in care may compromise care for people with special needs. If guidelines are misused by auditors, there is risk for inappropriately giving poor ratings to clinicians who purposefully deviate from recommendations for legitimate reasons. Reimbursement incentives and fear of legal ramifications when acting outside of recommendations may prompt physicians to adhere to recommendations even when a specific clinical situation may require a practice variation. Finally, even when recommendations are applicable to a given patient, the treatment may not represent the best use of limited health care resources.

Evidence of Effectiveness

Guidelines are able to improve care in theory, but whether they do in practice requires evaluation. Research into the effectiveness of guidelines remains limited and their effect is often difficult to separate from that of other research and initiatives in the same field. Two examples of guideline that have been associated with widespread and beneficial changes in practice are the NKF-KDOQI (National Kidney Foundation–Kidney Disease Outcome and Quality Initiative) guideline on vascular access and their guideline update on anemia management. The vascular access guideline was an important catalyst for a number of changes that have led to significant increases in the use of native arteriovenous fistulas, whereas the 2007 update on anemia management, along with clinical trial evidence from the year before, has been associated with significant changes in hemoglobin levels in the ensuing years. Whether these changes translate into improved patient outcomes (eg, fewer infections and longer and better lives) rests entirely on the evidence that underpins the relevant guideline.

Entities Providing Guidelines Relevant to Nephrologists

There are numerous entities producing guidelines relevant to clinical practice in nephrology, some of which are listed in Table 1. These include international and national nephrology entities, international and national disease-focused societies, and government-linked organizations. There also are several useful guideline databases that permit users to search for guidelines by topic or entity (Table 1).

Additional Readings

» Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess.* 2004;8(6):iii-iv, 1-72.
» Lenzer J. Why we can't trust clinical guidelines. *BMJ.* 2013;346:f3830. http://www.bmj.com/bmj/346/bmj.f3830.
» Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ.* 1999;318(7182):527-530.

METHODS FOR DEVELOPING GUIDELINES

The expected benefits of guidelines depend heavily upon the quality of the development process. This process includes topic selection; guideline group composition and group process; systematic review, synthesis, and grading of the evidence; moving from evidence to recommendations; reporting; and peer review.

Additional Readings

» Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical Practice Guidelines We Can Trust.* Washington DC: The National Academies Press; 2011. http://www.ncbi.nlm.nih.gov/books/NBK209539.
» Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med.* 2012;156(7):525-531.
» Schunemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ.* 2014;186(3):E123-E142.
» WHO. *WHO Handbook for Guideline Development.* Geneva, Switzerland: World Health Organization; 2015.

Topic Selection

Priority Setting

Developing guidelines is labor intensive and therefore costly. Although there is no empirical evidence to guide the choice of specific methods for prioritizing topics, the major guideline development bodies (including those in nephrology) take similar criteria into account:

1. Domain specificity: covering topics relevant to people with kidney diseases and their carers;
2. Importance: conditions affecting many people or with great burden of illness to an individual, families, communities, or society as a whole;
3. Need: perceived necessity of a guideline, as indicated by relevant stakeholders;
4. Cost: monetary cost per person of managing the health condition;
5. Variation in practice: as a proxy for clinical uncertainty over best practice;
6. Expected benefits on health outcomes from successful implementation of the guideline: potential

Table 1. Examples of Entities Producing Clinical Practice Guidelines Relevant to Nephrology

Guideline Body	Region	Website
	Nephrology	
CSN (Canadian Society of Nephrology) Committee for Clinical Practice Guidelines	Canada	www.csnscn.ca/committees/clinical-practice-guidelines/library
UK-RA (UK Renal Association)	United Kingdom	www.renal.org/guidelines/
KHA-CARI (Kidney Health Australia–Caring for Australasians With Renal Impairment)	Australia & New Zealand	www.cari.org.au/
NKF-KDOQI (National Kidney Foundation–Kidney Disease Outcomes Quality Initiative)	USA	www.kidney.org/professionals/KDOQI/guidelines_commentaries
ERBP (European Renal Best Practice)	Europe	www.european-renal-best-practice.org/
SLANH (Latin American Society of Nephrology and Hypertension)	Latin America	www.slanh.org/
KDIGO (Kidney Disease: Improving Global Outcomes)	Global	www.kdigo.org/
ISPD (International Society for Peritoneal Dialysis)	Global	http://ispd.org/ispd-guidelines/
	Diabetes	
CDA (Canadian Diabetes Association)	Canada	www.guidelines.diabetes.ca/
ADA (American Diabetes Association)	USA	www.professional.diabetes.org/ResourcesForProfessionals.aspx?cid=84160
AACE (American Association of Clinical Endocrinologists)	USA	www.aace.com/publications/algorithm
EASD (European Association for the Study of Diabetes)	Europe	www.easd.org/index.php?option=com_content&view=article&id=93&Itemid=508
IDF (International Diabetes Federation)	Global	www.idf.org/guidelines
	Hypertension	
ASH (American Society of Hypertension)	USA	www.ash-us.org/About-Hypertension/Hypertension-Guidelines.aspx
ESH (European Society of Hypertension)	Europe	www.eshonline.org/guidelines/arterial-hypertension/
ISH (International Society of Hypertension)	Global	www.ish-world.com/activities/guidelines.htm
	General	
NICE (National Institute for Health and Care Excellence)	England	www.nice.org.uk/guidancemenu/conditions-and-diseases/kidney-conditions
SIGN (Scottish Intercollegiate Guidelines Network)	Scotland	www.sign.ac.uk/guidelines/index.html
WHO (World Health Organization)	Global	www.who.int/publications/guidelines/en/
	Guideline Databases	
GIN (Guidelines International Network)	Global	www.g-i-n.net/
NGC (National Guideline Clearinghouse)	USA	www.guideline.gov/
NHMRC (National Health and Medical Research Council) guidelines portal	Australia	www.clinicalguidelines.gov.au/

to improve health care decision making for patients and providers, to improve mismatch between actual and appropriate care, and to enhance care by improving patients' quality of life, reduce avoidable morbidity, or reduce avoidable premature death;

7. Availability of evidence: clinical research suggesting management strategies can produce a difference in outcomes.

The relative importance of these criteria in topic prioritization is unclear. Most often, decisions are made implicitly. Nephrology guideline entities that use specific criteria to prioritize guideline topics do not provide written documentation of the information that informed the specific criteria or how this information influenced judgments on the relative priority of different topics. In a review of methods for priority setting commissioned by the World Health Organization, Oxman and colleagues (see "Additional Readings") suggested the following:

- Criteria for establishing priorities should be applied using a systematic and transparent process.
- Because data to inform judgements are often lacking, unmeasured factors should also be considered – explicitly and transparently.
- The process should include consultation with potential end users and other stakeholders, including the public, using well-constructed questions, and possibly using Delphi-like procedures.
- Groups that include stakeholders and people with relevant types of expertise should make decisions. Group processes should ensure full participation by all members of the group.
- The process used to select topics should be documented and open to inspection.

Updating Published Guidelines

As medical practice develops and advances, guidelines inevitably become outdated. Successful priority setting should reveal whether updating a previously published guideline takes preference over developing a new topic. Whether an update is needed will largely be determined by the extent to which new evidence has emerged that indicates that a change in the previously advocated practice strategies could lead to better outcomes or potentially reduce costs.

Guideline Adaptation and Appraisal

Adapting guidelines is a means used by some guideline entities of integrating one or more guidelines produced by other entities. Adaptation may be chosen as an alternative to developing a guideline anew and may reduce duplication of effort and increase efficiency of guideline entities working in the same clinical arena. Adaptation also may aid collaboration between discrete guideline entities, further enhancing efficiency. Adaptation might be desirable over simple endorsement, in which treatment recommendations need to change to suit jurisdiction-specific constraints in delivery of care. In their manual, the ADAPTE Collaboration detailed a systematic approach to identify, appraise, and endorse or modify recommendations to suit the local context. Every guideline is assessed for quality of development and reporting using the Appraisal of Guidelines Research & Evaluation (AGREE) II instrument. AGREE II is an internationally validated rigorously developed tool using 23 signaling questions to evaluate the quality of 6 domains and provides a means for making explicit judgments about guideline quality. The structure of AGREE II is shown in Table 2.

Additional Readings

» AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care.* 2003;12(1):18-23.

» Brouwers M, Kho ME, Browman GP, et al; on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *CMAJ.* 2010;182(18):E839-E842.

» Field MJ, ed. *Setting Priorities for Clinical Practice Guidelines.* Washington DC: National Academy of Sciences; 1995.

» Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. *Health Res Policy Syst.* 2006;4(1):14-21.

» The ADAPTE Collaboration. The Adapte Process: Resource Toolkit for Guideline Adaptation. Version 2.0. http://www.g-i-n.net/document-store/working-groups-documents/adaptation/adapte-resource-toolkit-guideline-adaptation-2-0.pdf.

» Vernooij R, Sanabria A, Sola I, Alonso-Coello P, Martinez Garcia L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. *Implement Sci.* 2014;9(1):3-12.

Table 2. Structure of AGREE II, an Instrument Used for Appraising the Quality of Clinical Practice Guidelines

Domain	Questions Within Domain
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described.
	2. The health question(s) covered by the guideline is (are) specifically described.
	3. The population (patients, public, etc) to whom the guideline is meant to apply is specifically described.
Stakeholder involvement	4. The guideline development group includes individuals from all relevant professional groups.
	5. The views and preferences of the target population (patients, public, etc) have been sought.
	6. The target users of the guideline are clearly defined.
Rigor of development	7. Systematic methods were used to search for evidence.
	8. The criteria for selecting the evidence are clearly described.
	9. The strengths and limitations of the body of evidence are clearly described.
	10. The methods for formulating the recommendations are clearly described.
	11. The health benefits, side effects, and risks have been considered in formulating the recommendations.
	12. There is an explicit link between the recommendations and the supporting evidence.
	13. The guideline has been externally reviewed by experts prior to its publication.
	14. A procedure for updating the guideline is provided.
Clarity of presentation	15. The recommendations are specific and unambiguous.
	16. The different options for management of the condition or health issue are clearly presented.
	17. Key recommendations are easily identifiable.
Applicability	18. The guideline describes facilitators and barriers to its application.
	19. The guideline provides advice and/or tools on how the recommendations can be put into practice.
	20. The potential resource implications of applying the recommendation have been considered.
	21. The guideline presents monitoring and/or auditing criteria.
Editorial independence	22. The views of the funding body have not influenced the content of the guideline.
	23. Competing interests of guideline development group members have been recorded and addressed.

Note: Adapted and abbreviated from http://www.agreetrust.org/, accessed March 30, 2015.
Abbreviation: AGREE, Appraisal of Guidelines Research and Evaluation.

Guideline Development and Group Composition and Processes

Group Composition

Guideline development is not a task that can be taken on by an individual. A multidisciplinary team is necessary for ensuring adequate consideration of the breadth of evidence and the different values attached to possible outcomes, as well as for securing ownership and support among target audiences. Ideally, teams represent the clinicians taking care of the patients (eg, nephrologists and clinicians in other specialties, primary care physicians, nurses, and paramedical staff) and other stakeholders potentially affected by guideline implementation (patients, their caregivers, and the policy makers determining resource use). It also is preferable that guideline groups include experts in epidemiology and statistics, guideline development, or implementation. However, the desire for wide representation needs to be balanced against timelines, costs for bringing collaborators together, and the challenges of effectively managing groups beyond 15 participants.

Conflicts of Interest

Guidelines should be based on scientific evidence, critical appraisal of the potential biases of that evidence, and objective clinical judgment in linking the evidence to what patients value. Financial or other ties to companies that are influenced by a guideline may negatively affect the panelist's ability to approach a scientific question objectively. Conflict of interest may not only bias the interpretation of evidence, but also determine which questions a guideline seeks to answer and unduly influence the process of generating recommendations. An example that became the source of heated debate is the 2006 KDOQI guideline for the management of anemia in people with chronic kidney disease. One recommendation therein proposed to increase hemoglobin target concentrations from 11-12 to 11-13 g/dL, although the available evidence indicated that a higher target was not associated with better survival or reduced risk for cardiovascular events. At the time, 14 of the 16 guideline development team members reported having received consultant fees or research funds from at least one of the companies that would potentially benefit from the guideline change. In addition, both the chair and cochair had financial relationships with the company that funded the guideline and stood to benefit from its recommendations.

Much less acknowledged, and possibly far more difficult to address, is intellectual conflict of interest. We often believe science to be a dispassionate pursuit of facts and that researchers willingly change their minds as new evidence emerges. In reality, human beings often find it extraordinarily difficult to look at their own results through the eyes of a detached observer or to change their prevailing beliefs. For most entities developing kidney guidelines, public disclosure of financial interests has become commonplace; however, intellectual interests are rarely considered and declared intellectual conflicts have limited consequences for guideline development participation. Many experts are skeptical that disclosure alone minimizes the bias. In its standards for developing trustworthy guidelines, the Institute of Medicine has advocated exclusion of those with conflicts of interest from leadership positions. The World Health Organization has adopted increasingly rigorous management strategies, excluding altogether those with conflicts of interest from the recommendation-generating process. Ideally, a guideline development panel would only include individuals without conflicting interests, whether financial or intellectual in nature. It would also include the most knowledgeable and those best able to convince others to implement the recommendations in practice. Unfortunately, well-known experts often have the most sweeping competing interests. Excluding them entirely from the process may not be possible without losing important insights into specific content or necessary support for effective implementation. The Institute of Medicine has recognized this: rather than demanding that all guideline development members be free from any competing interest, it has advocated keeping their participation to a minimum.

Group Processes

Although a multidisciplinary team is necessary for valid guideline development, this does not guarantee success. Successful guideline development relies on functioning group processes. What can go wrong? If one person dominates the discussion, resulting recommendations may diverge from the evidence or may not represent the opinion of the group; this pattern is known as minority influence. Likewise, groupthink may occur, a phenomenon in which a desire to reach agreement may override concerns about accuracy and drive people away from the evidence. Multidisciplinary groups are especially at risk here because established authorities often are the most active participants and may exert undue influence over other members of lower professional status.

Group process biases may be reduced by careful selection of the group leader, who should have the authority to enforce any ground rules, ensure complete group participation by providing equal opportunities for contribution from all members, give group members' arguments proper attention when articulating recommendations, maintain constructive dynamics, and check consensus attainment while

guarding his or her own neutrality. In the United Kingdom, the National Institute for Health and Care Excellence (NICE; a guidelines entity) seeks to appoint guideline working group chairs from nontopic specialties: "The Chair is appointed for their expertise and skill in chairing groups, and although they may have some knowledge of the topic, this is not their primary role in the group. Specialist knowledge is provided by other Committee members..." (www.nice.org.uk/article/pmg20/chapter/decision-making-committees).

More formal group process methods include established consensus building strategies such as nominal group techniques and Delphi methods, which tend to surpass informal methods in achieving agreement. The nominal group technique is structured group brainstorming that involves problem identification, solution generation, and decision making. The Delphi method is a widely used group communication process that aims to achieve a convergence of opinion from topic experts by 2 or more rounds of anonymous data gathering with subsequent statistical feedback.

Patient and Public Involvement

Quality standards increasingly stress the value of patient and public involvement in guidelines. Involving patients, their carers, or advocates is necessary to ensure that the views and experiences of these individuals are reflected and that the guideline is patient centered and covers issues and outcomes important to them. However, best methods of adequately soliciting consumer perspectives are not clear and also take time and money. If consumer perspectives are solicited without proper thought to the aims and methods required for achieving their integration, attempts to include such participation run the risk for remaining tokenistic, leaving consumers with a marginal role at best.

In 2012, G-I-N PUBLIC, a workgroup of the Guidelines International Network devoted to effective patient and public involvement in developing and implementing guidelines, identified 3 key strategies for patient and public involvement:

1. Consultation: aiming to collect views regarding needs, experiences, and expectations. This can occur at the scoping stage or when the draft guideline has been developed and can be targeted to relevant consumer groups and/or open to the public. Consultation may involve inviting comment on draft documents, center on systematic reviews of patient and public perspectives, and include surveys, focus groups, or individual interviews.
2. Participation: aiming to have active involvement in deliberation, to foster a collective perspective on guideline development, and to agree on common group decisions over guideline content. Although in theory participation can facilitate compromise between people with varying perspectives, adequate support is required for it to be effective.
3. Communication: aiming to provide patients and public with comprehensible information on diagnostic or treatment recommendations, enabling joint decision making based on the best available evidence.

For nephrology guidelines groups, the engagement of patients and the broader community with their work remains limited and in need of development.

Additional Readings

» Coyne DW. Influence of industry on renal guideline development. *Clin J Am Soc Nephrol.* 2007;2(1):3-7.
» Eccles M, Grimshaw J, Shekelle P, Schunemann H, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implement Sci.* 2012;7(1):60-67.
» G-I-N PUBLIC toolkit: Patient and Public Involvement in Guidelines. http://www.g-i-n.net/working-groups/gin-public/toolkit.
» Haller MC, van der Veer SN, Nagler EV, et al. A survey on the methodological processes and policies of renal guideline groups as a first step to harmonize renal guidelines. *Nephrol Dial Transplant.* 2015;30(70):1066-1074.
» Marshall E. When does intellectual passion become conflict of interest? *Science.* 1992;257(5070):620-623.
» Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess.* 1998;2(3):i-iv 1-88.
» Strippoli GFM, Tognoni G, Navaneethan SD, Nicolucci A, Craig JC. Haemoglobin targets: we were wrong, time to move on. *Lancet.* 2007;369(9559):346-350.

Systematic Literature Review/Development Processes

Scoping/Determining Key Questions and Outcomes

This is possibly the most important phase of the guideline development process, but team impetus to move forward to evidence review is sometimes at odds with spending time deliberating guideline questions and outcomes. Arriving at an evidence-based recommendation requires understanding which questions must be answered to get there, what evidence is available for consideration, and which outcomes need to be assessed. Typically, questions include identification of groups at risk for a health problem, diagnostic test accuracy, benefits and harms of different treatments, and significance of prognostic factors, among other issues. For example, to develop a recommendation for or against screening for cardiovascular disease in kidney transplant candidates, one would need to know if there are effective treatments to improve outcomes in candidates with severe asymptomatic coronary artery disease, if tests can accurately diagnose the condition, the potential harms

of testing, and who should be screened. Failure to articulate these key questions, and to define what evidence is admissible to answer them, can result in wasted time, effort, and money if it leads to gathering and analyzing data that are ultimately irrelevant to the recommendations in development.

Consider an example: if you tried answering a question on how to best screen for cardiovascular disease in kidney transplant recipients, how would you go about finding an answer? You could turn to your favorite internet browser and search the question verbatim. In Google, it would take 0.37 second to retrieve 179,000 citations (as of April 6, 2015). You would find guidelines and consensus statements and, if lucky, these would be at the top of your list and provide trustworthy advice. More likely is that these guidelines and consensus statements would be buried among many other entries, more or less relevant to the topic. There may be a few systematic reviews and hundreds of individual studies investigating the incidence of cardiovascular disease among kidney transplant candidates, the risk factors associated with having asymptomatic coronary artery disease, the diagnostic accuracy of various noninvasive tests in detecting disease, the association between asymptomatic coronary disease before and the outcomes after transplantation, and benefits and harms of treatments for asymptomatic disease before transplantation. There will be many more related to the same question but covering different patient groups, and there will be thousands of citations unrelated to the problem. Finally, you may miss crucial information because it is not picked up by the browser's search algorithm. This example illustrates that efficiently generating evidence-based advice requires understanding which questions need to be answered in what order, what evidence is required, and how to adopt a systematic strategy that will allow its retrieval.

Members of a guideline development team must unambiguously define which outcomes and which measurement time points they consider pivotal for decision making. This step forces the team to articulate in advance the extent to which particular outcomes need to be affected to support a particular recommendation. Outcomes must reflect what is important to patients. In the example of screening for cardiovascular disease, it is likely that patient-centered outcomes would include all-cause and cardiovascular death, as well as severe complications such as acute myocardial infarction. Important outcomes may also include indicators of physical or emotional well-being, such as functional status or related quality of life. Data for these health outcomes are often harder to ascertain because clinical trial investigators tend to measure surrogate outcomes to reduce the cost, sample size, and duration of trials. However, assumptions about the proximity of surrogates to clinically important outcomes are often not well informed. For example, it is tempting to assume that benefits of lipid-lowering treatment on total cholesterol level will result in downstream reduction in cardiovascular death, but this may not be the case for people with end-stage kidney disease.

Framing Questions for Systematic Review: PICO

Individual key questions must be framed such that relevant evidence can be systematically searched and selected. Framing is facilitated by addressing each part of the mnemonic PICO:

- P (patient and/or problem): How would you define a group of patients akin to the ones of interest? What are key common characteristics of these people?
- I (intervention or indicator): Which main interventions are you contemplating? What diagnostic test or exposure?
- C (comparison): What is the main alternative with which you wish to compare? Are you trying to decide between 2 medications, a medication and no treatment, or 2 diagnostic tests?
- O (outcomes): What is/are the outcome(s) of most interest in this clinical setting? Are clinicians or patients seeking benefit or reduced adverse events?

Table 3 shows how the clinical question on screening for coronary artery disease in kidney transplant candidates can be divided in sequential specific questions and converted into a focused format that can be answered more readily by research evidence.

Systematic Searching and Study Selection

Information sources. When precise guideline questions have been framed into an answerable format, the next step is to determine where to best look for an answer. Medical research data are searched most effectively through electronic databases, which offer citations and abstracts of journal articles and books. There are a wide variety of such databases with overlapping, complementary, or unique content areas, and each has its strengths and weaknesses; the "best" choice will largely depend on the type of question being asked and the optimal study design to answer it. It is beyond the scope of this article to cover all options so we have limited discussion to the databases most immediately useful to guideline developers: The Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE), the Cochrane Central Register of Controlled Trials (CENTRAL), the Medical Literature Analysis and Retrieval System (MEDLINE), and the Excerpta Medica Database

Table 3. Converting a Clinical Question to Framed Guideline Questions Using the PICO Format

Specific Question	Rationale	Requirement	PICO Framework
What is the incidence of cardiovascular mortality in kidney transplant candidates?	If cardiovascular mortality was nonexistent, there would be no rationale for screening for it	The condition should be an important health problem	Background question; found in textbooks, registry reports
What is the association between severe coronary artery disease and future risk for cardiovascular mortality in kidney transplant candidates?	If there is no relation between coronary artery disease and cardiovascular death, there would no rationale to treat it	There should be a latent stage of the disease. The natural history of the disease should be adequately understood	P: kidney transplant candidates I: coronary artery disease C: no coronary artery disease O: death
In transplant candidates, is treatment of asymptomatic coronary artery disease effective at reducing cardiovascular mortality before and after transplantation?	If there is no treatment available for which the benefits outweigh the harms, detecting the disease may not be useful	There should be a treatment for the condition. There should be an agreed policy on whom to treat	P: kidney transplant candidates with asymptomatic coronary artery disease I: percutaneous coronary intervention, CABG, medical treatment C: no treatment O: death
In transplant candidates, what is the diagnostic test accuracy of noninvasive investigations for detecting asymptomatic coronary artery disease?	If there is no test available that can detect the disease with reasonable accuracy, there is no rationale for screening. If the test is unacceptable for patients, attempting to screen will not work	There should be a test or examination for the condition. The test should be acceptable to the population	P: kidney transplant candidates I: resting electrocardiography, exercise stress electrocardiography, dobutamine stress echocardiography, myocardial perfusion scintigraphy C: coronary angiography O: coronary artery disease

Note: The clinical question in this example is "What is the best way to screen kidney transplant candidates for coronary artery disease?"

Abbreviations: CABG, coronary artery bypass grafting; PICO, patient/problem, intervention/indicator, comparison, outcomes.

(EMBASE) platform. Table 4 shows the databases most useful to search for the clinical questions developed earlier in this article, organized by the optimal study design for each question.

Search strategy structure. Executing an effective search to identify all studies relevant to the specific question at hand is a critical and often challenging step in the guideline development process. In identifying the totality of evidence for a question, one must recognize that the quantity of published and unpublished material is vast, much research remains unpublished, there are limitations of indexing systems within electronic databases and in the scope of individual databases, sensitive searching comes at the cost of retrieving irrelevant studies while specific searches may miss important evidence, and individuals invested in the content area may believe they already know all the relevant research within a particular field.

A typical search strategy consists of a logical combination of terms referring to the population; the intervention, index test, or exposure; possibly the comparator; the outcomes; and finally the study design. For each component, relevant search terms include a combination of keywords that database providers use to index the publications and documents listed within their database. These keywords make up the controlled vocabulary thesaurus (eg, Medical Subject Headings [MeSH] for MEDLINE and the Cochrane Library; EMTREE for EMBASE) and allow simultaneous retrieval of similar articles without specifying each synonym and spelling variation used by authors. Because indexing is at times subjective and incomplete, systematic search strategies typically include free text words (in title and abstract) to ensure important citations are not missed. Table 5 shows an example of a framed question translated into a sensitive search strategy.

Systematic Evidence Review Including Bias Appraisal of Included Studies

What Is Bias and Why Do We Need to Consider It?

Bias, a systematic error or deviation from the truth occurring in results or inferences, can result in over- or underestimation of the causal relationship between an intervention, an exposure, or a diagnostic test and an

Table 4. Framed Questions, Recommended Study Designs, and Databases to Search

Question	Study Design	Database[a]	Content
What is the association between severe coronary artery disease and future risk for cardiovascular mortality in kidney transplant candidates?	Systematic reviews of cohort studies, cohort studies	MEDLINE, EMBASE	Systematic reviews, RCTs, cohort studies, case-control studies, case reports, diagnostic cross-sectional studies, editorials, etc
In transplant candidates, is treatment of asymptomatic coronary artery disease effective at reducing cardiovascular mortality before and after transplantation?	Systematic reviews of RCTs	CDSR	Cochrane systematic reviews of RCTs
		DARE[b]	Non-Cochrane systematic reviews of RCTs
	RCTs	CENTRAL	RCTs
		MEDLINE, EMBASE	Systematic reviews, RCTs, cohort studies, case-control studies, case reports, diagnostic cross-sectional studies, editorial, etc
In transplant candidates, what is the diagnostic test accuracy of noninvasive investigations for detecting asymptomatic coronary artery disease?	Systematic reviews of cross-sectional diagnostic studies, cross-sectional diagnostic studies	MEDLINE, EMBASE	Systematic reviews, RCTs, cohort studies, case-control studies, case reports, diagnostic cross-sectional studies, editorials, etc

Abbreviations: CDSR, Cochrane Database of Systematic Reviews; DARE, Database of Abstracts of Reviews of Effects; CENTRAL, Cochrane Central Register of Controlled Trials; EMBASE, Excerpta Medica Database; MEDLINE, Medical Literature Analysis and Retrieval System; RCT, randomized controlled trial.

[a]All databases are available through Ovid. In addition, MEDLINE is available through PubMed; EMBASE via Elsevier; CDSR, DARE, and CENTRAL via The Cochrane Library.

[b]No longer updated as of April 2015.

outcome. Bias results from factors that may be known or unknown that provide a possible alternative explanation for the result attributed to the relationship between the intervention of interest and the outcome. Such factors have 3 features: they affect the outcome (eg, age), are unequally distributed between the experimental and control group, and do not represent an intermediate step in the causal chain between the exposure and outcome. In practice, the extent to which biases affect the results cannot be known; however, there is substantial evidence that specific shortcomings in the design, conduct, or analysis of studies result in bias. Some study designs are inherently more vulnerable to bias than others. For example, a randomized controlled trial (RCT) differs from an observational cohort study in that randomization aims to protect against selection bias while the blinding of participants and personnel attempts to minimize performance bias. However, not all RCTs produce equally trustworthy results because not all are conducted and reported optimally. RCTs produce treatment effect estimates that are exaggerated by up to 11% if the sequence generation is inadequate or unclear, 7% if the allocation concealment is inadequate or unclear, and 13% if blinding of participants and researchers does not occur. Estimates of exaggeration are greatest for meta-analyses assessing subjective outcome measures. Because each type of study (eg, RCT vs cohort study) has different opportunities for introducing bias, any assessment tool aiming to distinguish studies providing more or less trustworthy results needs to be tailored to the particular study design under evaluation.

Tools for Assessing Risk of Bias in Studies

Risk of bias can be used interchangeably with internal validity, or the degree to which a study's design and conduct were likely to have prevented biases. Many tools have been developed to aid in the assessment of potential bias in studies of different designs, each with their own strengths and weaknesses. Table 6 lists a suggested bias assessment tool for each study design. Without exception, these tools are the result of consensus between statisticians, epidemiologists, and review authors and are based on systematic review of empirical evidence.

Appraising a Body of Evidence for Individual Outcomes Across Studies

Regardless of the primary aim of the guideline literature review (eg, comparing interventions, assessing the accuracy of a test, or estimating the average effect of a risk factor), the next step is to examine the totality of evidence for each outcome across studies and draw conclusions about the summary effect. The degree of confidence in these conclusions will depend on several factors; the

Table 5. Anatomy of a Search Strategy: In Transplant Candidates, Is Treatment of Asymptomatic Coronary Artery Disease Effective at Reducing Cardiovascular Mortality Before and After Transplantation?

PICOM Component	Target	Search Syntax	Explanation
P	Kidney transplant candidates	**1:** MeSH descriptor: [Kidney Transplantation] this term only **2:** (kidney or renal) adj transplant*.ti,ab,kw **3:** #1 OR #2	**1:** Controlled vocabulary term to index articles about kidney transplantation; will include articles in which authors use "renal transplantation," "renal grafting," etc, instead of kidney transplantation. **2:** Addition of text words "kidney transplantation" and "renal transplantation" as a safety net to pick up articles that were indexed incompletely or incorrectly. "adj" used as Boolean operator to indicate the words "kidney" or "renal" need to be adjacent to "transplantation" in title, abstract, or listed keywords. **3:** Boolean operator "OR" indicating inclusion of all citations indexed as kidney transplantation plus all citations picked up by line #2, including all with both.
	With asymptomatic coronary artery disease	**4:** MeSH descriptor: [Cardiovascular Diseases] this term only **5:** MeSH descriptor: [Myocardial Ischemia] explode all trees **6:** coronar*.ti,ab,kw **7:** #4 OR #5 OR #6 **8:** #3 AND #7	**5:** Controlled vocabulary term to index articles about myocardial ischemia. "Explode all trees" is an option allowing inclusion of all terms hierarchically positioned underneath "Myocardial Ischemia," eg, "Angina Pectoris" and "Coronary Diseases." **8:** Boolean operator "AND" indicating inclusion of all citations picked up by line 3 and line 7 at the same time, ie, the intersection of both searches.
I	Percutaneous coronary intervention	**9:** MeSH descriptor: [Percutaneous Coronary Intervention] explode all trees **10:** PCI*.ti,ab,kw **11:** #9 OR #10	**10:** "*" is a truncation symbol; it retrieves any number of characters after the word stem or no characters.
	Coronary artery bypass grafting	**12:** MeSH descriptor: [Coronary Artery Bypass] explode all trees **13:** CABG*.ti,ab,kw **14:** #12 or #13	
C	Medical treatment	–	If the comparator is no specific treatment or standard care, which would also be given to the "interventional" group, then this is usually not included in the strategy to allow adequate sensitivity.
O	Death	–	For questions concerning interventions, outcome terms are usually not included to allow adequate sensitivity.
M	Systematic review of RCTs	**1:** Review.pt AND Medline.tw **2:** Meta analysis.pt **3:** (systematic* AND (review* OR overview*)).tw **4:** metaanaly*.tw **5:** meta-analy*.tw **6:** OR/1-5	In The Cochrane Library, CDSR, DARE, and CENTRAL are searched simultaneously; because CDSR and DARE preferentially include systematic reviews, methodological filters are not required.

(Continued)

Table 5 (Cont'd). Anatomy of a Search Strategy: In Transplant Candidates, Is Treatment of Asymptomatic Coronary Artery Disease Effective at Reducing Cardiovascular Mortality Before and After Transplantation?

PICOM Component	Target	Search Syntax	Explanation
M (cont'd)	RCTs	**1:** randomized controlled trial.pt **2:** controlled clinical trial.pt **3:** randomized.ab **4:** placebo.ab **5:** clinical trials as topic/ **6:** randomly.ab **7:** trial.ti **8:** OR/1-7 **9:** animals/ NOT (humans/ AND animals/) **10:** 8 NOT 9	Because CENTRAL preferentially includes RCTs, methodological filters are not required when searching The Cochrane Library. **1**: The limiter ".pt" refers to publication type. Articles indexed as being an RCT are searched. **3**: Articles with the word "randomized" in the abstract are searched. **7**: Articles with the word "trial" in the title are searched. **10**: The Boolean operator "NOT" indicates articles selected in line 9 are excluded. Lines 9 and 10 together exclude studies exclusively conducted in animals.

Note: Search development is built around the PICOM anatomy of the question posed, and is built for searching MEDLINE via OvidSP. Similar syntax can be used to search The Cochrane Library (see notes in last column), or when searching MEDLINE via PubMed (see help materials at the National Library of Medicine website). The M in PICOM refers to methods. Abbreviations: CDSR, Cochrane Database of Systematic Reviews; CENTRAL, Cochrane Central Register of Controlled Trials; DARE, Database of Abstracts of Reviews of Effects; MEDLINE, the Medical Literature Analysis and Retrieval System; MeSH, Medical Subject Heading; RCT, randomized controlled trial.

reliability of results of the individual studies is not the only determinant. Other features include evaluating the consistency and precision of the results across studies, applicability of results to the targeted population, and whether publication bias is likely.

Consistency of evidence relates to heterogeneity of findings across studies. Study results can vary due to differences in key aspects of the design (eg, trials with lower or higher risk of bias) or of the study participants (eg, those treated vs those not yet treated with dialysis), interventions (eg, doses, duration of treatment, and co-interventions), and outcomes (eg, definitions and measurement time point). If important variability among studies remains unexplained, it will reduce the confidence in any inferences. However, assessing the level of variability that is regarded as "important" is a matter of judgment.

Directness of evidence, also known as applicability or external validity, is also important to consider when appraising evidence. In the case of guideline development, the applicability of evidence can be reduced by differences between study participants and the population for whom the guideline is being developed. This might be driven by difference in baseline risk, culture, lifestyle, delivery of care, or the availability of technologies and resources. Directness is also compromised if measured study outcomes differ from those deemed critically important by the guideline development team (eg, Kt/V vs health-related quality of life). Further, directness is lessened if a study uses surrogate outcomes instead of important objective patient outcomes (eg, creatinine clearance vs the need for kidney replacement therapy).

Precision relates to confidence in the magnitude of an estimated effect. Precision is usually represented by the 95% confidence interval, which can be interpreted as the range within which the effect plausibly lies. The confidence interval needs to be sufficiently narrow and it must encompass the minimal effect considered clinically important.

Publication bias occurs if "negative studies" remain unpublished or unidentified. When only published studies are considered, interventions appear more effective, tests appear more accurate, and risk factors appear more important. It is difficult to formally assess for publication bias outside of the systematic review setting, and statistical methods all have their limitations.

The GRADE (Grading of Recommendations Assessment, Development and Evaluation) working group provides a framework to facilitate formal evaluation of the reliability of a body of evidence (Table 7).

Table 6. Tools for Assessing Risk of Bias/Critical Appraisal Tools

Study Design	Tool	Bias	Mechanism
Randomized controlled trials	Cochrane risk of bias tool for randomized controlled trials	Selection bias	Bias due to inadequate random sequence generation
		Selection bias	Bias due to inadequate allocation concealment
		Performance bias	Bias due to inadequate blinding of participants and researchers
		Detection bias	Bias due to inadequate blinding of outcome assessment
		Attrition bias	Bias due to incomplete outcome data
		Reporting bias	Bias due to selective reporting
		Other bias	Bias due to other mechanisms
Nonrandomized studies of interventions	A Cochrane risk of bias assessment tool: for nonrandomized studies of interventions (ACROBAT: NRSI)	Confounding bias	Bias due to confounding
		Selection bias	Bias in selection of participants into the study
		Misclassification bias	Bias in measurement of interventions
		Performance bias	Bias due to departures from intended interventions
		Attrition bias	Bias due to missing data
		Detection bias	Bias in measurement of outcomes
		Reporting bias	Bias in selection of the reported result
Cohort studies	Newcastle-Ottawa quality assessment scale: cohort studies (NOS: Cohort)	Selection bias	Representativeness of exposed cohort
		Selection bias	Selection of the nonexposed cohort
		Misclassification bias	Ascertainment of exposure
		Detection bias	Demonstration that outcome was not present at start of study
		Selection bias	Comparability of cohorts on basis of the design and analysis
		Detection bias	Assessment of outcome
		Detection bias	Adequacy length of follow-up for outcome occurrence
		Attrition bias	Adequacy of follow-up of cohorts
Case-control studies	Newcastle-Ottawa quality assessment scale: case-control studies (NOS: Case-control)	Selection bias	Adequacy of case definition
		Selection bias	Representativeness of cases
		Selection bias	Selection of controls
		Selection bias	Definition of controls
		Selection bias	Comparability of cases and controls on the basis of the design or analysis
		Detection bias	Ascertainment of exposure
		Detection bias	Same method of ascertainment for cases and controls
		Attrition bias	Nonresponse rate
Cross-sectional diagnostic test accuracy studies	Quality assessment of studies of diagnostic accuracy (QUADAS-II)	Selection bias	Bias in selection of participants into the study
		Spectrum bias	Applicability of test results to target population of interest
		Performance bias	Bias in conduct or interpretation of the index test (blinding)
		Detection Bias	Applicability of index test, its conduct, or interpretation to review question
		Performance bias	Bias in conduct or interpretation of reference standard
		Performance Bias	Applicability of target condition as defined by reference standard to review question
		Attrition bias	Bias through patient flow
Systematic review	Assessing the methodological quality of systematic reviews (AMSTAR)	Reporting bias	Was an 'a priori' design provided?
		Selection bias	Was there duplicate study selection and data extraction?
		Selection bias	Was a comprehensive literature search performed?
		Publication bias	Was the status of publication (ie, grey literature) used as an inclusion criterion?
		Publication bias	Was a list of studies (included and excluded) provided?
		Selection bias	Were the characteristics of the included studies provided?
		Information bias	Was the scientific quality of the included studies assessed and documented?
		Information bias	Was the scientific quality of the included studies used appropriately in formulating conclusions?
		Detection bias	Were the methods used to combine the findings of studies appropriate?
		Publication bias	Was the likelihood of publication bias assessed?
		Conflict of interest bias	Was the conflict of interest included?

Additional Readings

» Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011;64(4):395-400.

» Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol.* 2011;64(4): 380-382.

Table 7. GRADE Method of Rating the Quality of the Evidence

Step 1: Starting Grade According to Study Design	Step 2: Lower if	Step 3: Higher if	Step 4: Determine Final Grade for Quality of Evidence
Randomized trials = High Observational studies = Low	**Risk of bias** −1 Serious −2 Very Serious **Inconsistency** −1 Serious −2 Very Serious **Indirectness** −1 Serious −2 Very Serious **Imprecision** −1 Serious −2 Very Serious **Publication bias** −1 Likely −2 Very Likely	**Large effect** +1 Large +2 Very Large **Dose response** +1 Evidence of a gradient **All plausible confounding** +1 Would reduce a demonstrated effect +1 Would suggest a spurious effect when results show no effect	High (4 plus: ⊕⊕⊕⊕) Moderate (3 plus: ⊕⊕⊕◯) Low (2 plus: ⊕⊕◯◯) Very Low (1 plus: ⊕◯◯◯)

Adapted from Balshem et al (GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64[4]:401-406) with permission of Elsevier.

Note: GRADE specifies 4 categories for the quality of a body of evidence: from high to very low. The quality of evidence for an outcome is initially rated high if it originates predominantly from randomized controlled trials and low if it originates from observational data. These ratings can be subsequently downgraded or upgraded if any of a number of additional criteria are met.

Abbreviation: GRADE, Grading of Recommendations Assessment, Development and Evaluation.

» Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.

» Scottish Intercollegiate Guidelines Network (SIGN). Chapter 5: assessing the quality of evidence. In: *SIGN 50: A Guideline Developer's Handbook*. Edinburgh, Scotland: SIGN; 2014: 89-106. http://www.sign.ac.uk/pdf/sign50.pdf.

» Shea B, Grimshaw J, Wells G, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7(1):10-16.

» Sterne JAC, Higgins JPT, Reeves BC, et al. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT- NRSI), Version 1.0.0. http://www.riskofbias.info.

» Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

» Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.

GENERATING RECOMMENDATIONS

Rating Strength of Recommendations and Quality of the Evidence Across Key Outcomes

A recommendation is an individual statement proposing the best course of action for a given situation. Recommendations can be for or against a strategy and can have varying degrees of conviction. The strength of a recommendation is established by the quality of the evidence across key outcomes, by judgments related to the balance between the desirable and undesirable consequences of a strategy, by variability in values and preferences, and by the associated costs. Clear communication of both the intended message and the value judgments that influence the recommendation generating process is key.

Guideline developers apply formal approaches for separately rating the strength of recommendations and the overall quality of the underlying evidence, resulting in the assignment of categorical scores. Currently, GRADE is the most widely used grading system and is used by all major nephrology guidelines organizations. It categorizes a recommendation as "strong" if benefits clearly outweigh harms or vice versa, the confidence in the evidence for important health outcomes is high, health care consumers agree on the value of different outcomes, and the proposed strategy represents a wise use of resources. It categorizes a recommendation as "weak" if the trade-off between benefits and harms is less certain, the quality of evidence is low, there is high-quality evidence suggesting that wanted and unwanted effects are essentially in balance, or health care consumers have diverging ideas on which outcomes matter most to them. Although the discrete approach to rating recommendations can be criticized for categorizing what in reality is a continuum, it has the benefit not only of simplicity but also of direct association with actions on the part of clinicians, patients, and policy makers. To communicate the overall quality of the evidence supporting a recommendation, a summary rating (ranging from "high" to "very low") is assigned; this represents

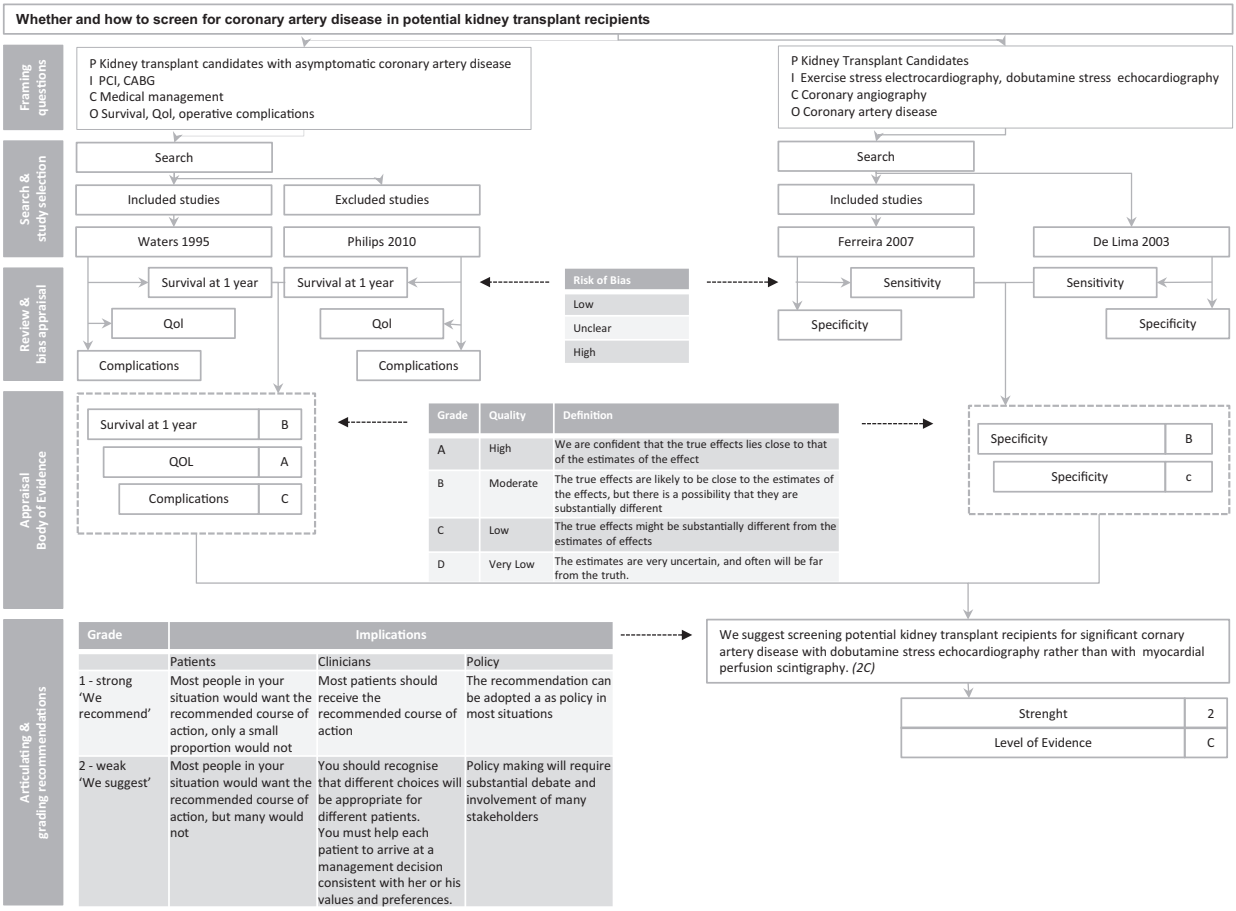

Figure 1. From framing the question to generating the recommendation, a worked example of whether and how to screen for coronary artery disease in potential kidney transplant recipients. Abbreviations: CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention; QOL, quality of life.

the lowest quality of evidence for any of the outcomes that are critical to decision making (Table 7). Figure 1 provides a worked example showing how the GRADE approach is used in practice.

Articulating Recommendations

It is reasonable to assume that specific wording affects how statements are interpreted and whether they can be implemented consistently by clinicians. Most guideline organizations use the specific terminology proposed by GRADE to confer the strength of a recommendation: "we recommend" for a strong statement and "we suggest" for a weak statement. This standardization of language and methods is a positive feature of guidelines because it increases consistency in how recommendations are interpreted. However, recommendations have often been criticized for being nonspecific and failing to indicate precisely what should be done under a given set of conditions. Vague recommendations using structures such as "patients with condition A should receive treatment B if considered necessary" are usually not very useful in clinical practice because they are dependent upon the interpretation of what constitutes "necessary." A well-worded recommendation should specify the population and the specific conditions under which it applies. Unless it is obvious, the recommendation should also specify the comparator. For example, in the statement "in patients at increased risk of contrast-induced acute kidney injury, we recommend intravenous volume expansion with isotonic saline or sodium bicarbonate," the strength of this recommendation may differ depending on whether the alternative is oral hydration or no hydration at all.

Additional Readings

» Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol.* 2013;66(7):726-735.
» Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol.* 2012;66(2):151-157.

Reporting and Peer Review

Reporting refers to the level of detail contained within the guideline document and how a guideline

will be made public (eg, print or online). Peer review refers to how the guideline document will be reviewed before its publication and how it can be assessed (eg, for errors), both internally and externally, by stakeholders who were not members of the guideline development team.

Reporting a Guideline

The user makes the ultimate ruling on a guideline's trustworthiness. An informed decision requires detailed description of the methods used for developing the guideline, including all the items covered in this Core Curriculum. Moreover, in reporting a guideline, it is essential to provide a narrative rationale describing the link between every recommendation and its supporting evidence so that the user can judge the extent to which a recommendation can be justified. The narrative should cover both anticipated benefits and potential risks that may arise when the guideline recommendations are implemented. In 2003, a conference on guideline standardization proposed a list of 18 core items to report for a guideline. The major organizations producing guidelines for patients with kidney diseases currently use similar formats, covering the majority of the expected items.

External Review

A guideline development group is limited in the knowledge it can bring to the table and the perspectives it can offer. External review helps ensure balance, comprehensiveness, accuracy of the scientific evidence, and validity of the rationale underlying the recommendations. It also allows for feedback on the clarity and practicality of recommendations and contributes to the engagement of stakeholders. There are several types of external review, serving different purposes and requiring input from different people.

Reviewers can be specifically invited, usually for their perceived ability to contribute. They may include leading clinical researchers or key opinion leaders and may be asked to identify missing research and mistakes in the description of studies, interpretation of their quality, or reporting of their results. In addition, involving expert peer reviewers may increase a sense of ownership and support of those who are most influential in the field of interest. Invited reviewers may include methodologists, who are recruited to identify possible biases, asymmetry in the presentation of evidence, or flaws in the logic applied when translating evidence into recommendations. Practicing clinicians in relevant fields also may be asked to review to identify any ambiguities in the recommendations that may come to light when they are applied in practice or any effects on workflow that may hamper implementation. Other peer reviewers may include representatives from federal agencies, professional organizations, specialty societies, health care providers, peer-reviewed journals, and guideline panels authoring related guidelines. They may include representatives from advocacy organizations, community groups, consumers, and public health organizations whose constituents may be affected by the guideline, as well as representatives from businesses and industry (eg, pharmaceutical or device manufacturers) affected by guideline recommendations.

A draft guideline also may be posted online for public consultation. In this case, the guidelines are typically made available for a defined period and are open for comment by any interested party. Public consultation provides an opportunity to create a sense of partnership, which may increase acceptability of the guideline as a whole and help promote awareness and adoption of the recommendations.

Finally, guideline developers can choose to include peer review through publication in a peer-reviewed journal. This may help establish the scientific credentials of the guideline and may facilitate its implementation through outreach to the readership of the journal. Disadvantages include delays due to the peer-review process and external control over guideline release dates. Another possible drawback is the potential for journal editors to exert undue influence over the guideline content, insisting on changes based on comments of a few reviewers that may or may not be substantiated by the evidence and agreed methodology.

Regardless of the type of external review, methods for dealing with criticism should comply with development methodology, and criticism can never simply be ignored. Groups need to adopt a system for recording, discussing, and processing elicited comments. Given that resistance to guideline implementation often comes from insufficient understanding of the methodology used, care needs to be taken that responses to external review are communicated adequately.

Additional Readings

» Shekelle P, Woolf S, Grimshaw JM, Schunemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. *Implement Sci.* 2012;7(1): 62-68.

» Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the conference on guideline standardization. *Ann Intern Med.* 2003;139(6): 493-498.

IMPLEMENTATION OF GUIDELINES

Producing a guideline does not change practice. The next step is guideline implementation. Detailed exposition on implementation science is beyond the scope of this article, but a brief overview is warranted. The aphorism that "guidelines don't implement themselves" poses a challenge for guideline developers because implementation requires a different set of skills and resources from those required to develop the guideline. Implementation of a guideline requires that clinicians are aware of the guideline, understand the guideline and its implications for their service and patient care, and consider it important enough to tailor their practice to more closely reflect the guideline recommendations. Within this voyage from evidence summation to implementation, there are many potential barriers at the various layers of the health care delivery system; therefore, it is not surprising that most guideline development groups concede the responsibility of implementation to health service delivery organizations.

Dissemination

The easiest part of guideline implementation is that of dissemination. There are numerous portals through which guidelines can be publicized, with most kidney guideline groups publishing recommendations in peer-reviewed journals, as well as on their websites (Table 1). Many guideline groups also publicize their products using the more usual tools that researchers use (eg, in the proceedings of scientific meetings) in addition to promoting documents though local specialty societies. Furthermore, with the increasing availability of mobile internet devices, smartphone apps that summarize clinical guidelines are becoming more widely used (eg, KDIGO and NICE). However, dissemination alone is not a very effective tool for driving practice change.

Additional Implementation Challenges

Beyond informing clinicians about the guideline, the more effective guideline implementation tools share common ground with health care quality improvement and assurance methods. These methodologies include audit and feedback, barrier analyses, and use of key opinion leaders and extend to the use of reminder systems and multifaceted interventions delivered at multiple levels of the health care system. The most common approach used to address guideline–practice gaps in health care is the Plan-Do-Study-Act approach (which is designed to be applied in iterative cycles) to generate and analyze data from a health care process, use results to develop interventions to alter that process, and measure the outcomes of these interventions before further action is taken (Fig 2). However, development of new implementation methods and evaluation of existing strategies are ongoing and likely to evolve further in coming years. Clearly these methods require different skill sets from those needed for guideline development, as well as significant investment in the form of staff time and project infrastructure. Organizations that fund and manage health systems are likely to be in the best position to drive such processes, but may also be the groups least likely to be aware of the guideline evidence.

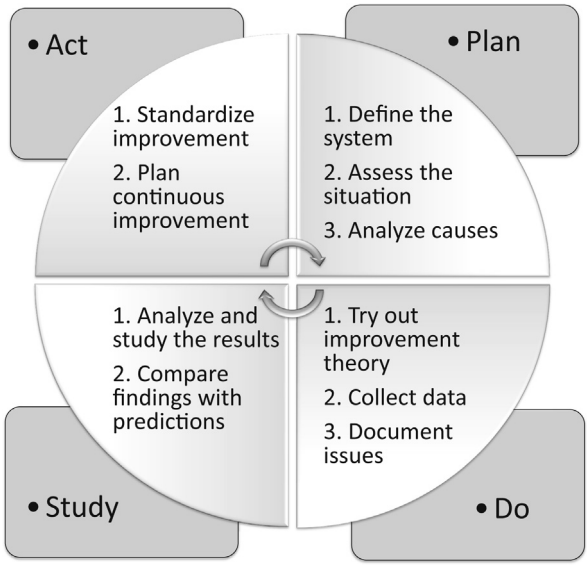

Figure 2. The Plan-Do-Study-Act (PDSA) cycle for guideline implementation at the clinical level.

Additional Readings

» Coleman J, Katz E, Menzel H. The diffusion of an innovation among physicians. *Sociometry.* 1957;20(4):253-270.
» Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ.* 1997;157(4):408-16.
» Gallagher MP, Cass A, Craig JC. Applying evidence into routine clinical care at a unit level: the exemplar of renal anaemia management. *Nephrology.* 2010;15(4):429-433.
» Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet.* 2003;362(9391):1225-1230.

ACKNOWLEDGEMENTS

Dr Webster is Deputy Co-ordinating Editor of Cochrane Kidney and Transplant. Dr Nagler is a member of the ERBP methodology support team.

Support: None.

Financial Disclosures: The KHA-CARI Guidelines Steering Committee, of which Dr Gallagher is chair, previously received funding support from pharmaceutical companies. The other authors declare that they have no relevant financial interests.