

---


Electronic Theses and Dissertations, 2004-2019

---

2010

## Clinical Practice Guidelines For Emerging Ultrasound Applications Drafting For Validity And Usability

Kathi Keaton Borok  
*University of Central Florida*

 Part of the [Technical and Professional Writing Commons](#)  
Find similar works at: <https://stars.library.ucf.edu/etd>  
University of Central Florida Libraries <http://library.ucf.edu>

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### STARS Citation

Borok, Kathi Keaton, "Clinical Practice Guidelines For Emerging Ultrasound Applications Drafting For Validity And Usability" (2010). *Electronic Theses and Dissertations, 2004-2019*. 1595.  
<https://stars.library.ucf.edu/etd/1595>

CLINICAL PRACTICE GUIDELINES FOR EMERGING ULTRASOUND APPLICATIONS:  
DRAFTING FOR VALIDITY AND USABILITY

by

KATHI KEATON BOROK  
B.S. Webster University, 2002

A thesis submitted in partial fulfillment of the requirements  
for the degree of Master of Arts  
in the Department of English  
in the College of Arts and Humanities  
at the University of Central Florida  
Orlando, Florida

Fall Term  
2010

© 2010 Kathi Keaton Borok

## **ABSTRACT**

Clinical practice guidelines (CPGs) are viewed by many people with interests in health care as valuable tools for reducing practice variations that undermine patient outcomes and increase medical costs. However, guidelines themselves vary in quality. Assessment tools generally base quality measures on strength of guidelines' evidence base, but particularly for newly emerging applications of ultrasound, standards for measuring guideline quality are controversial. The validity of a guideline is considered likely when strong research-based evidence supports its recommendations, but for newer medical procedures such as emerging ultrasound applications, available evidence is sparse. Existing assessment tools must be modified if they are effectively to measure the validity of these guidelines built on immature evidence.

Focusing on ways document drafting affects CPG validity, this study rated six guidelines using the Appraisal of Guidelines Research and Evaluation (AGREE) tool which was customized according to categories of guideline purposes and their differing features of validity. Fine-tuning AGREE in this way may create a more consistent, informative method of evaluating guidelines for emerging applications, and standards established in such an instrument may be useful as a template during the guideline development process.

Results from my analyses illuminate several common omissions that weakened documents. Most guidelines did not describe an updating procedure or identify areas for future research, but results also highlighted some highly effective techniques for building validity. Notable examples include providing full credentials for expert drafters, and embedding statement references directly in the text. From the results of the analysis, I conclude that, although the

adapted assessment tool I used needs additional adjustment, it may refine analysis of guidelines for emerging ultrasound guidelines and conversely serve as a useful tool during their development process.

## ACKNOWLEDGMENTS

It has been a privilege and joy to work with the members of my thesis committee, Drs. Dan Jones, Melody Bowdon, and Lori Brodtkin. Responsibility for the contents of the paper is entirely mine, but I am indebted to these three gifted professors for their contributions to its development process.

It is my honor to especially thank Dr. Dan Jones for his invaluable guidance, support, and technical direction throughout. A consummate professional as editor and advisor, Dr. Jones is also a generous, caring mentor. Without his calm, steady reassurance, it would not have been possible for me to complete this project.

I am also most grateful to Dr. Melody Bowdon. Dr. Bowdon honed my technical skills in the classroom and provided valuable advice on my thesis, but, more importantly, she touched my life and helped me find my voice as a writer.

To Dr. Lori Brodtkin I extend warmest thanks. Dr. Brodtkin brought talent, enthusiasm, and interest to working with me on my thesis, and helped sustain me with her gentle encouragement during the days just before my defense.

I wish also to thank several colleagues whose contributions shaped my work. For the inspiration they provided for the topic, I am indebted to Drs. Joshua Copel and Steven Goldstein, and to my dear friend Paula Woletz. My warm thanks also go to Dr. Carmine Valente for providing frequent doses of encouragement, humor, and understanding, and to Dr. David Paushter for sending articles my way and indulging me in long, patient discussions of guidelines.

Completing this thesis would not have been possible without the support of friends and loved ones. With most tender gratitude I thank Imogen for warning me about the slippery

noodles, David for editing me, and both of them for their unflagging encouragement. It is with pleasure that I also offer special thanks to the Graham family for inviting me into the splashing inspiration of their eastern shore paradise where many pages of this thesis were conceived.

Finally, I thank my precious, wonderful daughters. All the way down the long and winding path to this work's completion they patiently supported and encouraged me, and believed in me. With deepest love and gratitude, I dedicate this thesis to Terra, Harmony, Jenny, and Mackenzie.

## TABLE OF CONTENTS

TABLE OF CONTENTS.....	vii
LIST OF FIGURES .....	x
LIST OF TABLES.....	xi
CHAPTER ONE: INTRODUCTION.....	1
Quest for Effective Medical Care .....	3
Practice Variation and Patient Outcomes .....	4
Rising Medical Costs .....	7
History of CPGs.....	7
The Role of CPGs in Reducing Practice Variation.....	8
Introduction to Ultrasound Guidelines and Emerging Applications.....	10
CHAPTER TWO: LITERATURE REVIEW.....	15
The Rhetorical Challenge .....	16
Guideline Attributes.....	21
Characterizing the Evidence .....	23
Presenting Recommendations.....	27
Emerging Users and Appropriateness.....	30
CHAPTER THREE: METHODOLOGY .....	33
The AGREE instrument.....	33
Domain 1: Scope and Purpose .....	34
Domain 2: Stakeholder Involvement .....	35
Domain 3: Rigor of Development.....	36



Domain 4: Clarity of Presentation .....	37
Domain 5: Applicability .....	38
Domain 6: Editorial Independence .....	38
Rubric Development .....	39
Rubric Version 1 (RV1).....	39
Guideline categories: Qualifying the rubric.....	41
Guideline Category Assessment .....	42
Category Group 1: Stratified Recommendations .....	43
Category Group 2: Multi-option Recommendations .....	43
Category Group 3: Performance Recommendations .....	44
Adjusting the Rubric for Category Differences .....	45
CHAPTER FOUR: ANALYSIS.....	50
Guideline Assessments .....	50
Guideline 1: American College of Emergency Physicians.....	50
Guideline 2. American College of Emergency Physicians.....	56
Guideline 3: American Urological Association.....	63
Guideline 4: American College of Cardiology Foundation, et al .....	67
Guideline 5: American Institute of Ultrasound in Medicine .....	76
Guideline 6: American Institute of Ultrasound in Medicine .....	80
Summary of Findings.....	84
CHAPTER FIVE: CONCLUSIONS .....	89
How should guidelines for emerging ultrasound applications be assessed?.....	89
Is it appropriate to develop guidelines for these applications? .....	92

If so, how can these guidelines best be constructed for validity and usability? .....	92
APPENDIX A: ACCREDITATION .....	96
APPENDIX B: CPG QUALITY ASSESSMENT AND CONSTRUCTION TOOLKIT .....	98
LIST OF REFERENCES .....	102

## LIST OF FIGURES

Figure 1: Rubric Version 1 (RV1) .....	40
Figure 2: Category-Adjusted Guideline Assessment Tool .....	49
Figure 3: CPG Development Template.....	95

## LIST OF TABLES

Table 1: Category Relevance .....	46
Table 2: Category-Adjusted Guideline Quality Assessment .....	48
Table 3: Clinical Policy: Critical Issues in the Initial Evaluation and Management of Patients Presenting to the Emergency Department in Early Pregnancy (2003) .....	52
Table 4: Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients With Suspected Appendicitis (2010).....	58
Table 5: Consensus Statement on Urologic Ultrasound Utilization (2007) .....	64
Table 6: Appropriateness Criteria for Stress Echocardiography (2008).....	70
Table 7: Practice Guideline for the Performance of the Musculoskeletal Ultrasound Examination (2007).....	77
Table 8: Practice Guideline for the Performance of Fetal Echocardiography (2010).....	81

## **CHAPTER ONE: INTRODUCTION**

During the last several years the practice of medicine has come under intense scrutiny, both from external entities and from within the medical community itself. While the rapidly advancing science of medicine has produced tremendous diagnostic and therapeutic tools, the cost of medical care has spiraled out of control, with no end in sight. Moreover, patient safety at the hands of medical caregivers has been called into question. Concerns about these two issues – patient safety and medical costs – have precipitated a major reappraisal of the medical system and delivery of health care in the United States.

How can we provide care that is better and more cost-effective? Fundamental to this question is the problem of unwarranted practice variation leading to inconsistent care and a spectrum of practices for any given application ranging from best, most effective to ineffective and even dangerous. Basing care decisions on “best practices” as defined in clinical practice guidelines (CPGs) is viewed by many as a solution, but for some applications there are multiple and differing guidelines, and for some emerging applications there are as yet no guidelines at all (Carter 1649).

My thesis examines clinical practice guidelines, with particular focus on how construction of guideline documents affects overall guideline quality. To establish context for discussing written communication issues, I explore the role of CPGs in promoting best clinical practices, how the quality of existing guidelines can be assessed, and how future guidelines might be developed and constructed to reduce unwarranted practice variation and exert maximum positive impact on patient outcomes and cost of care. My thesis examines the role of

guidelines in promoting best clinical practices, how the quality of existing guidelines can be assessed, and how future guidelines might be developed and constructed to reduce unwarranted practice variation and exert maximum positive impact on patient outcomes and cost of care. In the sections that follow I describe the patient safety and cost issues associated with practice variation, and I discuss how those problems precipitated a proliferation of CPGs. I define the distinguishing characteristics of these documents and examine methods of assessing CPG quality. Finally, I explore special demands associated with developing guidelines for emerging applications. This background information will lay a foundation for construction of a guideline development template the American Institute of Ultrasound in Medicine, along with its collaborative partners, can use to produce useful, high quality guidelines for emerging applications.

I propose that three major challenges are likely to present themselves for developers of emerging applications guidelines. First, there is the question of whether, as opponents will charge, it is even appropriate to devise guidelines for narrow exams to be performed by clinicians who are not ultrasound specialists. And indeed the question of appropriateness should be addressed in such a guideline. The second challenge will be overcoming potential philosophical discord among a diverse group of practitioners practicing in a variety of settings with a variety of skill sets. Third, for novel applications, there will be very little evidence on which to base guideline recommendations. Each of these theoretical issues presents challenges for content experts and for technical communicators drafting the documents.

In my thesis I propose to discuss these guideline needs of emerging ultrasound users' groups, and to explore ways the document drafters can contribute to the development of sound guidelines for this diverse group. These are my key questions:

- 1) How should guidelines for emerging ultrasound applications be assessed?
- 2) Is it appropriate to develop guidelines for these applications?
- 3) If so, how can these guidelines best be constructed for validity and usability?

### **Quest for Effective Medical Care**

The history of medicine recounts the eternal human quest for ways to combat disease and promote health and well-being. But disease processes are complex, and the quest for most effective, or “best,” diagnostic and treatment approaches has led medical investigators down myriad fruitless, even bizarre, pathways. Consider the practice of medicine in colonial America. Colony “doctors,” most of whom were educated entirely through apprenticeships, attacked their patients’ ails with copious blood-letting and all manner of strange and vile concoctions. In the 1600s, “best practice” in the treatment of apparent blood diseases meant a prescription for tincture of charred toad. For madness there was St John’s wort, and for itching, patients were prescribed a brandy-turpentine-hog’s fat-fish worm cocktail. From what may have been the only medical text book available in the Colonies, its author, Dr. Frances Packard, advised “For paines in ye Brest or Limmes; Weare a Wilde Catts skin on ye place grieved” (20).

One might suspect that patients’ safety in early America was threatened by proximity to the very treatments doctors regarded as “best practices.” In fact, some historians suggest that during the 1793 Yellow Fever epidemic, more people died from medical “care” than from the disease itself.

For the first century of America’s history, colonial physicians had little more than lore and superstition on which to base their patient care decisions. “Evidence-based medicine” as we define it today was essentially nonexistent in those early years of American medicine. But in

1765 the quest to unravel the mysteries of disease sprouted academic roots with the opening of North America's first medical school at the College of Philadelphia, now the University of Pennsylvania. King's College, predecessor to Columbia University, quickly followed suit, conferring its first medical degree in 1770. In 1823 the University of Maryland built the first US teaching hospital, providing the fledgling American medical community with a formal research home where investigators could examine theories and grow evidence. Evidence of most effective practices presumably provides clinicians with clear answers to how to best care for patients.

### **Practice Variation and Patient Outcomes**

Unique insights are essential to the art of medicine. In fact, some of medicine's most important advances have proceeded from individual flashes of intuition, only later supported by scientific evidence. An example is the germ theory of disease. In 1847, well before the germ theory was confirmed, Hungarian physician Ignac Semmelweis noted the much lower rate of "childbed fever" among women delivered by midwives, rather than by physicians who attended them with unwashed hands, immediately after performing autopsies. Based on his clinical observations and intuitive conclusions, Semmelweis established a "guideline" requiring physicians to wash their hands before deliveries, and in so doing, greatly reduced maternal morbidity and mortality. It was not until 1890 when the germ theory of disease was finally scientifically confirmed by Robert Koch.

For the 19<sup>th</sup> and most of the 20<sup>th</sup> century, the practice of medicine relied more heavily on the common sense and powers of observation of the care giver than on formal evidence and broadly endorsed practice guidelines. Some of us still remember the family doctors of the 1950s



and 60s. Equipped with their black bags, physicians with their arcane knowledge were trusted by their patients as much for their sage judgment as for their medical educations. In fact, from formal education doctors learned about treatment options, but it was from their experiences that they built what were very personal “evidence bases” on which to found care decisions. The doctor who saw one patient improve while following a particular regimen might generalize and apply that technique to all patients with the same complaint; a colleague in another community, informed by different experiences, might perceive “best” patient outcomes in response to another regimen. The practice of medicine varied among physicians according to their individual experiences and interpretations.

Such variations in the practice of medicine seem charmingly Norman Rockwell, but, in fact, these disparities are associated with inconsistent, often poor, quality of care. According to Charlton, even with access to 21<sup>st</sup> century medicine, some clinicians “persist in old-fashioned and eccentric practices under the guise of ‘clinical freedom’” (Charlton 99-101). Many other complex factors are associated with inconsistent care as well. Geographic variations have been reported, particularly with respect to prescription of drugs, and disparities in care have been reported associated with numerous social factors, including patients’ ethnicity, health provider, economic status, and gender. Schuster et al reviewed 48 articles that assessed the quality of preventive, acute, and chronic health care services in the United States. From their review these investigators concluded that among all populations “there are large gaps between the care people should receive and the care they do receive” (Schuster 517-563).

Clearly “wrong,” or “inappropriate” care harms patients as occurs in the “wrong site/wrong procedure/wrong patient” scenario. “Too much” care in the form of unnecessary medications and services is also injurious, increasing the patient’s risks and side effects. A rising

number of unnecessary caesarean sections is increasing both maternal risk from surgical complications and fetal risk from sometimes-preterm delivery is an example of too much care.

As conventional wisdom would predict, increased morbidity and mortality result from “too little” care, as when potentially beneficial diagnostic or therapeutic services are not offered. According to a Dartmouth Atlas Project Topic Brief on effective care, more than 68,000 people die each year from causes related to hypertension, but fewer than 65% of hypertensive patients receive recommended care. Fewer than two thirds of older Americans receive the pneumonia vaccine, resulting in nearly 100,000 preventable deaths each year. According to the Dartmouth report, “effective care is underused largely because our health care system lacks the means of supporting systematic compliance with treatment guidelines” (Dartmouth ).

Nearly two centuries after the birth of medical education in the US, we shake our heads and smile at the 18<sup>th</sup> century colonial doctor’s leeches and poisons. But even with our more sophisticated knowledge, pharmacologies, and technologies, the practice of medicine remains imperfect. Clinicians still seek answers to questions as to how to provide best care for their patients, and patients still suffer at the hands of medical care givers. In fact, in 1991 the *New England Journal of Medicine* reported that, according to statistics gathered by a Harvard Medical Practice Study, an estimated 33,000-98,000 people die each year as a result of medical errors (10). That disturbing report catalyzed a widesweeping movement in the medical community to search out root causes for errors, to find ways to safeguard patients, and to optimize delivery of patient care. CPGs have been cited as key to more consistent care and better patient outcomes.

## **Rising Medical Costs**

Reducing unwarranted practice variation is driving a strong movement towards CPGs as tools for promoting patient welfare; another factor driving the guidelines movement is their hoped-for potential to improve the medical cost/benefit ratio and help contain the unsustainable rate of increase in health care costs. From \$294.3 billion in 1969, in 2009 the cost of health care skyrocketed to \$2.5 trillion, and higher spending levels are not associated with improved outcomes (Woolhandler 768). The Commonwealth Fund annually compares healthcare costs and outcomes of several countries. In its 2007 report, among Australia, Canada, France, Germany, Japan, Norway, Sweden, the United Kingdom, and the United States, in the US per capita health care costs at nearly 17 percent of the gross domestic product (GDP) were the highest while life expectancy and infant mortality were the worst. By reducing practice variation, can CPGs lead to better patient outcomes and more efficient use of medical resources?

## **History of CPGs**

Guidelines have been used in medicine for centuries. Perhaps the earliest example in Western literature is the medical ethics guideline ascribed to Hippocrates. Thought to have been written around 400 BC, the “Hippocratic Oath” defined an ideal “Do no harm” practitioner-patient relationship still relevant today.

The modern guidelines movement began more than 50 years ago, and surged late in the 20<sup>th</sup> century as concern grew regarding the rising cost of health care. According to Farmer, the rate of healthcare spending began accelerating in the 1960s when Medicaid/Medicare began offering extremely low cost care to poor and senior citizens. HMOs were created to try to regain control, but these managed care organizations failed to effectively slow spending.

Answers to the growing crisis were sought in outcomes research, and in 1987 William Roper, Chief of the Health Care Financing Administration (HCFA) launched a series of seminars focused on improving health care. One conclusion from Roper's initiative was reported by researcher John Wennberg who demonstrated variations in healthcare, and healthcare costs, stemming from the way doctors were doing business. Also as a result of the initiative, Robert Brook of the Rand Corporation developed criteria for judging appropriate performance of medical procedures. In 1988 Roper's effectiveness initiative was formalized by HCFA, and the National Institutes of Health (NIH) and the National Center for Health Services Research (NCHSR) joined as partners to continue the outcomes research.

It was in 1989 that guidelines became a real national focus. That year Congress charged the NCHSR, eventually renamed the Agency for Healthcare Research and Quality (AHRQ), with leading in the task of developing and evaluating CPGs. Interest in evidence-based guidelines was evolving, and in 1989 the Institute of Medicine (IOM) launched an initiative to evaluate the quality of CPGs and plan ways to advance the art of guideline development. CPGs were formalized as a genre in 1990 when the IOM formally defined them as "systematically developed statements to assist practitioners and patients in choosing appropriate health care for specific clinical conditions" (Field 38).

### **The Role of CPGs in Reducing Practice Variation**

According to a report from the Agency for Healthcare Research and Quality (AHRQ), a research arm of the US Department of Health and Human Services, "Scientific studies have

proven that following best practice guidelines reduces suffering and patient mortality, while improving quality of life and clinical outcomes” (??).

Guyatt et al acknowledge that practice should be guided by findings reported in the literature, but they maintain that for many practitioners the medical literature is inaccessible. “Researchers studying the literature must accurately interpret a diverse range of journal articles and analyze the overall statistical findings, as well as the circumstances unique to each included study” (from AHRP website <http://www.ahrq.gov/clinic/epc/qgapfact.htm>).

Compounding the difficulty for individuals attempting to stay abreast of the literature, its volume is staggering, and increasing. According to Guyatt, “every few years the literature doubles in size, and every year we seem to have less time to weigh it,<sup>1</sup>” (Guyatt or Durack – check this reference - ) Especially as technological innovation explodes and research into application outcomes flourishes, it becomes increasingly difficult for clinicians to review all available literature on a topic.

As long ago as 1951 Garland identified the problem for radiologists: “The expanding literature on radiology threatens to engulf the serious worker. Since radiology embraces the diagnostic and therapeutic applications of radiant energy in the entire field of medicine, this expansion is both understandable and inevitable” (??).

Garland’s 1951 predictions were correct. In fact, American diagnostic radiology researchers published 468 papers in 1960 and nearly 3000 in 1984 (Chew 1055); between 1990 and 2005 more than 7500 original articles were published in the journal of the Radiological Society of North America alone (Arrive 330). Rapidly accumulating medical research data is, indeed, expanding the literature to such a volume that clinicians can no longer rely on personal reading to keep them abreast of all research findings relevant to their areas of practice. To further

complicate the clinician’s task of digesting the literature, conclusions of different studies often conflict.”

One broadly supported solution to the problems with the literature (2 Goolsby) is to look to teams of experts to review all possible care options, weigh the evidence for and against each, and, based on the evidence, develop evidence-based care recommendations communicated through clinical practice guidelines (CPGs), Battista 7: “CPGs offer the potential to reduce uncertainty attributable to suboptimal decision-making due to unfamiliarity with relevant scientific knowledge” (provide source info at end of passage).

Across the board, the health care industry is focused on standards and guidelines as tools for promoting patient welfare. For example, some commercial health plan programs, or “3<sup>rd</sup> party payers,” have instituted strict quality measures of “privileging” and “site accreditation” as conditions for reimbursement for imaging services. DC Levin and VM Rao describe Blue Cross Blue Shield programs, and they report how those quality requirements have affected the delivery of health care. According to these authors, by reimbursing for only those services provided by practitioners with specific training, third party payers have helped reduce the number of studies completed by low-quality providers; by requiring site accreditation as a condition for payment, these commercial health plan providers are raising the bar on facility quality standards.

At issue: do CPGs work? How do guideline developers write these documents to be effective? And how can potential users assess the likeliness of reliability of recommendations?

### **Introduction to Ultrasound Guidelines and Emerging Applications**

In ultrasound, as in other areas of medicine, practice guidelines are developed and regularly revised to articulate the community’s agreed-upon minimum standards of care.

Currently the American Institute of Ultrasound in Medicine (AIUM) publishes 24 practice guidelines developed by its Clinical Standards Committee (CSC), most guidelines in collaboration with one or more other professional organizations. Writing and revising guidelines even for fairly standard practices is complicated and strongly rhetorical. Guidelines must be carefully worded to protect patients, protect doctors and other caregivers, satisfy regulating agencies like the Joint Commission, and satisfy third party payers. Complex collective voice, complex audience, complex purpose—these and other elements add up to complex content development and technical communication challenges.

While many medical guidelines address specific disorders, most current ultrasound guidelines outline performance and training standards for imaging a particular body structure, or group of structures. Traditionally, ultrasound examinations are performed in designated ultrasound laboratories by ultrasound imaging specialists. Outpatients travel to these facilities, and patients already receiving care in a facility are moved to the ultrasound lab for the exam. But the practice of ultrasound is changing; traditional body part imaging is no longer the only application for this very user-friendly technology, and the ultrasound department with its specialized clinicians is no longer its only setting. Because ultrasound is free of ionizing radiation, highly portable, and relatively inexpensive, medical caregivers from specialties outside traditional imaging boundaries are picking up transducers and trying them out for an expanding array of new uses, many at the “point of care” in labor and intensive care units, emergency departments, and surgical suites. Emergency physicians examining critically injured patients are using ultrasound for a quick look for internal bleeding; nurses are using it to guide insertion of PICC lines; anesthesiologists are looking with ultrasound to localize a nerve for injection. And

that short list covers only a few of the possible new uses. Traditional imaging ultrasound is being joined by an explosion of emerging applications.

The characteristic that distinguishes an application as “emerging” generally falls into one of these categories: use of ultrasound to examine structures not traditionally evaluated with ultrasound; or use of the technology to perform traditional imaging tasks in nontraditional settings, by practitioners whose primary specialty is not ultrasound imaging, or with a more limited scope. Some characteristics of these emerging applications are controversial and may pose challenges to achieving consensual guidelines. Of the features that characterize emerging applications, the two that may prove most contentious relate to exam scope and practitioner provider. To protect patients from inconsistent, insufficient care, the standards of the traditional sonography community require performance of "complete" examinations, and only by credentialed sonographers or physicians as described by existing CPGs. Any guideline verbiage that is perceived to dilute these standards will likely meet with strong opposition.

Two major factors are driving interest in development of guidelines for emerging users of ultrasound. First, the number of new users is large and growing, and both users and patients would likely benefit from guidelines outlining the safest, most effective techniques for their new applications. Second, the current medical climate and the huge push for improved patient safety is promoting guidelines and standards as a means to reduce errors and improve effectiveness of medical care decisions. Exploring the task of developing practice guidelines for emerging users of ultrasound is timely and important both for practitioners and for their patients.

In Chapter Two, I review the literature review that describes the format and attributes that characterize clinical practice guidelines. I explore the challenges inherent in writing high quality



guidelines with a focus on the special challenges associated with emerging point-of-care ultrasound applications.

In Chapter Three, “Methodology,” I discuss the Appraisal of Guidelines Research in Evaluation (AGREE) instrument as a tool for assessing writing practices that affect the validity and usability of a document. AGREE is a validated and well accepted tool, but it is not entirely appropriate for assessing guidelines for emerging ultrasound applications or for other guidelines of similar purpose. Adjusting AGREE according to needs I assigned to emerging users’ guidelines, I developed an assessment tool to use for these documents.

In Chapter Four I present my analyses of six clinical practice guidelines that address various ultrasound applications. Using the rubric developed through the process described in Chapter Three, I assess text components of these documents and rate how effectively each one promotes guideline validity and usability. Although I assign a score to attributes and to each document overall, the primary purpose of my assessments is to identify writing practices that weaken guidelines and to highlight practices that make the documents strong. With this study I evaluate features and writing challenges fundamental to CPGs, but I direct special focus toward those attributes that may prove most important, and most challenging, for developers of guidelines for emerging ultrasound applications.

In Chapter Five, I describe conclusions based on the literature and on my experiment with guideline document assessment. Applying the conclusions I reached with respect to assessments, I developed a template for drafting emerging users’ guidelines. Finally, combining the assessment tool described in Chapter Three with the template described in Chapter Five, I provide a CPG Quality Assessment and Construction Toolkit (Appendix B). I believe findings

and conclusions from this thesis may be used by guideline drafters to enhance the quality of CPGs for emerging users.

## CHAPTER TWO: LITERATURE REVIEW

The rapid increase in the number of CPGs published during the past several years gives testament to people's belief in their potential to reduce practice variation. But guideline development as a discipline is scrutinizing itself. In their seminal 1993 article, Grimshaw and Russell state that "the quality of published guidelines is variable" (*Developing* 243). As a result of this concern, there is growing interest in ways to assess guidelines and to apply quality measures prospectively during the development process. In this literature review, I discuss the basic format and attributes that characterize CPGs, and I explore the demands inherent in writing high quality guidelines, with a focus on the special challenges associated with emerging point-of-care ultrasound applications.

Many CPG development groups have looked at ways to evaluate guideline quality. Some groups focus on topics related to gathering and grading evidence; others examine the method used to formulate recommendations; still others discuss strategies for implementation. I reviewed those topics, but with a focus on the role of the writer in effectively conveying the nature of those processes and findings. Next, I reviewed the literature that differentiates guidelines according to their purposes, and examined how their different functions might affect document characteristics. I gave special consideration to the guideline needs for emerging ultrasound applications.

Using University of Central Florida online library resources, I first sought relevant sources through a Medline search using these key words and key word phrases: clinical practice guidelines, practice guidelines, medical guidelines, guidelines, performance guidelines, history of guidelines, validity, collaboration, collaborative techniques. Using the same key words and

phrases, I searched the Internet using Google and Google Scholar. Some additional sources were identified from among bibliographic references in selected articles, and others were recommended by a physician colleague. To identify primary guideline documents for analysis, I searched the Internet using key words “medical professional organizations,” and then selected and downloaded CPGs from organizations’ websites. An e-mail from a patient is included with her permission.

### **The Rhetorical Challenge**

Compelling considerations underlie the writing challenges inherent in this specialized genre of medical literature, and recommendations conveyed by guidelines bear significant clinical, professional, and economic impact. Health outcomes are most important, but the cost of suggested care must be taken into account when advocating care practices, as well as the obligations guideline recommendations impose upon clinician’s resources. In a way, CPGs represent a negotiated understanding between all members of the audience, and the audiences for these highly consequential documents are multiple and complex. As an example, the primary audience for an obstetrical ultrasound guideline includes sonographers and sonologists who perform the exam according to the guideline-recommended protocol. Then to this already-diverse group, with its range of background experiences, add the physicians, physician assistants, and midwives who refer patients for ultrasounds based on indications listed in guidelines, and the more distant but still real audience of patients, insurers, legislators, and others.

Each of these reader groups views CPGs through a lens of special interest. For example, according to Grimshaw, the costs of recommended care may be of particular concern to third party payers. In fact, for major purchasers of health care, like the Centers for Medicare and

Medicaid Services (CMS), cost effectiveness is a critical issue. With the cost of medical care rising at a rate that exceeds inflation and the gross domestic product, these CPG purchasers will likely examine guidelines for evidence of clinical and cost effectiveness of recommended care.

Across the board the health care industry is focused on standards and guidelines as tools for promoting patient welfare. For example, some commercial health plan programs, or “third party payers,” only reimburse for imaging services performed by providers who are identified through accreditation programs of professional organizations like the American Institute of Ultrasound in Medicine (AIUM) as adhering to high standards of patient care. Since to be designated as “accredited,” clinical groups must agree to practice according to specific guidelines, CPG preferences of these large insurers bear significant implications for health care providers.

There is evidence of positive outcomes from insurers’ support of guideline-based practices. Levin and Rao describe Blue Cross Blue Shield programs, reporting how their quality requirements have affected the delivery of health care. According to these authors, by reimbursing for only those services delivered by practitioners with specific training, third party payers have helped reduce the number of studies completed by low-quality providers (535). By requiring site accreditation and adherence to particular guidelines as a condition for payment, these commercial health plan providers are raising the bar on facility quality standards. And movement toward accreditation and encouraging practice according to guidelines as a condition for compensation will continue. United Healthcare (UHC), one of the largest private insurance companies in the US, declares on its website that it now reimburses only accredited providers for performance of Computed Tomography, Magnetic Resonance Imaging, and Positron Emission Tomography, and some cardiology imaging procedures, and they are considering “future

application to other ultrasound services.” UHC’s website reports that by January 1, 2012, Medicare will also require provider accreditation as a condition for payment of some imaging fees.

Clinicians, insurers, and legislators have interests in reading CPGs, and some patients are reading medical literature, including practice guidelines, as well. Just a few years ago Americans regarded the principles of medicine as mysterious, largely inscrutable concepts; physicians with their secret knowledge were revered and rarely questioned. But today, thanks to the World Wide Web, a staggering volume of medical information (and misinformation) is readily available to anyone with access to the internet. Any lay person can google “H1N1” and reach roughly 132,000,000 relevant sites. The Internet surfer in a small town in Alaska need only type in keyword “headache” to learn about headache symptoms, diagnosis, and treatment from some 26,700,000 sites.

The cyber-information explosion is making lay diagnosticians of us all, and contributing to a paradigm shift in the clinician-patient relationship. For as Frederick Douglass learned, information access is empowering. In his autobiography, Douglass relates his master’s opposition to his wife’s teaching the young slave to read: “If you teach [Douglass] how to read, there would be no keeping him. It would forever unfit him to be a slave. . . . It would make him discontented and unhappy” (1071). As Douglass discovered, his master was right: information access leads to questioning.

Indeed, today’s more informed patients are much more likely than their grandparents to question their care providers, and 21<sup>st</sup> century patients’ outcomes expectations are greater. Americans witnessing the medical innovations and improved outcomes of the last few years have come to expect a lot from their medical care and care providers. In the following e-mail, a patient

recounts her family's difficult experience at the birth of her child, and conveys her frustration with failure of prenatal ultrasound clinicians to identify her baby's serious structural defect. The patient specifically references prenatal ultrasound guideline recommendations that were not followed during her exam.

2/26/2010

My son was born in December 2007 with Bladder Exstrophy. This condition went undetected prenatally by ultrasound. I am a nurse and have studied my son's ultrasounds (5 total), radiology reports, and fetal anatomy criteria to be met by ultrasound.

...

Our radiology report does not document umbilical cord insertion site, the route of the arteries to the umbilical cord, a definitive abdominal wall, emptying and filling of the bladder, or pelvic bone structure. The radiologist did document that the bladder could not be seen but later amended the report saying that the structure must be very small.

Our obstetrician never questioned these reports nor mentioned any concerns to me or my husband.

The only change to the outcome of our situation would have been preparation for birth, preparation financially, and the ability to shield our then 8 year old daughter from a frightening situation. Our son was born vaginally, in a rural hospital to staff who had never seen this condition, ill equipped to care for our baby. He was transported to Jacksonville, where we were questioned on several occasions as to why we did not receive prenatal care.

That day quickly turned from one of the most joyous days of my life to a nightmare. God has blessed us, and our son is doing great. He will have future surgeries and has endured four to date. My fear is that others will be faced with a similar situation, when it can be prevented.

Parents should have the opportunity to arrange for their birth to take place in a facility equipped to handle this type of case and the proper birthing method. They should have the opportunity to prepare themselves financially for this situation. We have incurred medical costs that we will not soon overcome. They should have the opportunity to learn about this birth defect and know what to expect after birth. This condition requires a progressive approach, babies will endure several surgeries, parents need to be trained in what they can do to help their child heal. They should have the opportunity to prepare their families, other siblings and grandparents, it is devastating to not know why no one can see or hold the long awaited bundle of joy.

My goal is to raise awareness in the medical community and the need for continuing education to keep technicians, radiologists and obstetricians current and informed in medical advances along with proper diagnosis given specific tests outcomes.

Clearly audiences for guidelines include a wide range of readers, from specialized clinicians to patients with no medical background, and these readers are often strongly vested in their perceptions of the recommendations and outcomes described in CPGs. To meet the rhetorical responsibilities imposed by these factors, developers of all types of clinical practice guidelines are charged with wording information accurately and in such a way as to be understood concordantly by all members of the audience. This is no simple task, for as Richards and Ogden note in “The Meaning of Meaning,” “Words... are a very imperfect means of communication” (1277). Particularly in light of potential ramifications of misinterpretation, it may not be sufficient for content experts to simply state the facts. As Kenneth Burke observed, “We can safely take it for granted that no one’s ‘personal equations’ are quite identical with anyone else’s” (1345).

Such ambiguity as Burke describes challenges interpretation of the most fundamental descriptor of clinical practice guidelines, the term that is often regarded as synonymous with high quality: “evidence-based.” McAlister et al observe that this seemingly explicit term may carry different meanings for guideline developers than for clinicians. To clinicians, “evidence-based” may seem to imply the high quality of RCTs while for guideline developers the term strictly means that recommendations are based on evidence that was derived from systematic search and evaluated as to strength. This important distinction illustrates the difficulty of writing meaningful, unambiguous, high quality guidelines.



## **Guideline Attributes**

In drafting these complex documents, technical communicators with their specialized wordsmithing skills can provide valuable assistance to the content expert members of guideline development committees. In fact, the overall quality of a guideline depends on both its ideas and its documentation. Grimshaw and Russell frame the broad concept of guideline development, intertwining the process of developing content with the task of drafting the final, implementable document. In the first in their series of articles, the authors identify and define nine fundamental attributes of CPG quality (*Developing Scientifically* 245):

- Validity
- Reproducibility
- Reliability
- Representative development
- Clinical applicability
- Clinical flexibility
- Clarity
- Meticulous documentation
- Scheduled review

Of these nine attributes described by Grimshaw, two directly relate to authors' technical task of document construction: *clarity* and *meticulous documentation*. According to Grimshaw and Russell, document clarity depends on accurate, unambiguous presentation of information in a format designed for easy access (*Developing Scientifically* 245). Meticulous documentation requires accurate reporting of all details of guideline development.

Two other attributes relate to the functional application of guideline recommendations: *clinical applicability* and *clinical flexibility*. The authors assert that guidelines have clinical applicability if the recommendations suit the defined patient population. The attribute of clinical flexibility provides leeway and guidance for clinical judgment to override recommendations. The attribute of scheduled review ensures that recommendations are based on current evidence and will be updated as evidence changes.

Grimshaw and Russell characterize “validity” as a more global attribute than the others, and of the most critical importance to guideline quality. It is also likely the attribute that will prove the most challenging for developers of emerging users’ guidelines to implement. So what does it mean for a guideline to be “valid”? How is that quality defined? According to Dorland’s Illustrated Medical Dictionary, validity is “the extent to which a measurement, test, or study measures what it purports to measure” (“Validity,” def. 1). The Agency for Healthcare Research and Quality (AHRQ) Glossary of Terms located on their website offers a similar definition, stating that validity is determined by “whether a test or technique actually measures what it is intended to measure.” The AHRQ applies that general definition with respect to guidelines, tying validity to expected health and cost outcomes. Similarly, in their 1995 article, Grimshaw et al specify that the validity of guidelines is measured by their effect on patient outcomes and resource utilization. They define guidelines as “valid if, when followed, they lead to improvement in patient outcome at acceptable costs” (*Implementing* 61).

According to Grimshaw and Russell , the quality of validity depends on AHRQ’s three remaining attributes of *reproducibility*, *reliability*, and *representative development*. They say that reproducibility means a guideline’s recommendations should seem appropriate to the evidence base as interpreted by various groups; reliability assures that groups other than the guideline

drafters will be likely to apply the guideline similarly and under similar circumstances. The authors assert that to promote these two conditions of reproducibility and reliability, the guideline development group must include *representatives of all relevant interests*.

Grimshaw's "attributes" inform the task of capturing validity in these documents. According to the AHRQ, three factors constitute the main influences on guideline validity: 1) the strength of the evidence on which recommendations are based, 2) the method used to evaluate the evidence, and 3) the relationship between the force of recommendations and the strength of the supporting evidence. In the next sections I review the literature related to each of these three influences, and I discuss the dual and complimentary roles of content experts and technical communicators in addressing these elements.

### **Characterizing the Evidence**

Various authors discuss factors that affect the quality and strength of evidence. Several authors stress the importance of a literature review as a basis for evidence quality. Drafters' meticulous reporting of the literature search and review process enhances the document's validity by communicating to users the degree to which they can rely on the guideline's recommendations. Through a formal, objective literature review, evidence relevant to a guideline's topic is located and considered. Grimshaw advises that guidelines developed "without a lit review may be biased towards reinforcing current practice rather than promoting evidence-based practice" (*Developing Scientifically* 246). In another article Grimshaw warns of the potential for bias when formal systematic literature reviews are absent from the development process and recommendations are based entirely on the literature familiar to and supplemented by the opinions of the committee members. According to Grimshaw, such recommendations

“often lag behind available evidence and may reflect individual enthusiasms rather than the knowledge base” (*Developing and Implementing* 61).

Shekelle et al provide details appropriate to this report item. In the second article of a series on guideline development, these authors outline an evidence search method, beginning with the step of consulting an up-to-date systematic review on the guideline topic. In the absence of an available systematic review, they describe using formal computer search strategies to locate relevant, appropriately designed studies.

The literature suggests that guideline drafters promote validity by accurately characterizing the strength of each piece of evidence that comprises a CPG’s evidence base. That strength of evidence, according to the Training Manual for The AGREE Collaboration Appraisal of Research and Evaluation (AGREE) Instrument, is related to “the extent to which all aspects of a study’s design and conduct can be shown to protect against systematic bias, non-systematic bias, and inferential error” (62).

Guideline drafters transparently report the risk of bias inherent in evidence by properly characterizing its study design and applicability. According to the ranking scheme of the AHRQ, evidence least vulnerable to bias is that which is obtained from systematic reviews and meta-analyses of randomized controlled trials. Since these studies include large numbers of participants, the effects of individual-specific, unintended variables are minimized, and proceeding results are highly reliable. Also highly reliable, but one step down because of lesser numbers, are randomized controlled trials. Vulnerability to bias increases for evidence from the next category of studies, non-randomized intervention studies, and increases even more with observational and non-experimental studies. The AHRQ ranks expert opinion as the lowest quality of evidence with the greatest risk of bias. Other guideline developers rank evidence

quality similarly, generally regarding the strength of evidence as inversely related to the possibility of study bias and directly related to the quality of the study design and implementation.

McAlister et al identify clinical relevance as another factor that affects CPG validity. According to McAlister et al, even guideline recommendations based on “high quality” evidence from internally valid RCTs are of less than “high” quality if clinically irrelevant or not directly applicable to the target population. These authors analyzed bibliographies of guidelines for treatment of type 2 diabetes mellitus. Of 15 “evidence-based” guidelines, there was very little concordance of reference selection, and the study cited most frequently among the guidelines’ references did not actually include patients with type 2 diabetes mellitus. To define the quality of their evidence bases more effectively, these authors advocate grading guidelines using schemes that take into account the clinical relevance and applicability of recommendations (e250).

Drafters’ meticulous reporting of the literature search and review process enhances the document’s validity by communicating to users the degree to which they can rely on the guideline’s recommendations. In a description of the process, Shekelle et al provide details appropriate to this report item. In the second article of a series on guideline development, these authors outline an evidence search method, beginning with the step of consulting an up-to-date systematic review on the guideline topic. In the absence of an available systematic review, they describe using formal computer search strategies to locate relevant, appropriately designed studies.

The authors describe how selected evidence may be rated, interpreted, and translated into recommendations of various strengths, depending on the quality of their evidence base. According to Shekelle et al “Recommendations based solely on clinical judgment and experience

are likely to be more susceptible to bias and self interest.” When expert opinion is incorporated, “the process needs to be made as explicit as possible” (Shekelle 595). Recommendation grades convey with what confidence guideline developers level of confidence that following the recommendation will lead to the predicted outcome.

Grimshaw and Russell discuss the relationship between the evidence and validity, and they rate the likelihood of various evidence bases leading to valid guidelines. From formal meta-analyses and graded systematic reviews, they predict “high” likelihood; with ungraded systematic reviews they predict “medium”; from unsystematic reviews or from expert opinion they predict “low” likelihood of guideline scientific validity. At issue in terms of bias and evidence synthesis: expert opinion alone tends to reinforce existing practices. Objective conclusions depend on unbiased consideration and interpretation of all possible solutions and outcomes (*Developing Scientifically* 245).

The tasks of grading the evidence according to its risk for bias, and rating recommendations according to assigned strength are simplified by many groups by symbolically encoding the information. Different organizations use varying codes, but as an example, following is the classification scheme reported by Shekelle et al (595):

#### Category of evidence

Ia-evidence for meta-analysis of randomised controlled trials

Ib-evidence from at least one randomized controlled trial

IIa-evidence from at least one controlled study without randomisation

IIb-evidence from at least one other type of quasi-experimental study

III-evidence from non-experimental descriptive studies, such as comparative studies, correlation studies, and case-control studies

IV-evidence from expert committee reports or opinions or clinical experience of respected authorities, or both.

### **Presenting Recommendations**

#### Strength of recommendations

A-directly based on category I evidence

B-directly based on category II evidence or extrapolated recommendation from category I evidence

C-directly based on category III evidence or extrapolated recommendation from category I or II evidence

D-directly based on category IV evidence or extrapolated recommendation from category I, II or III evidence

Inevitably evidence is viewed through the lens of cultural values, which helps explain variation in recommendations among different guidelines developed for a single application. For example, looking at cholesterol management guidelines, Battista et al note differences in the scopes between those of the United States and Canada. These authors contend that the difference reflects “a cultural preference for greater caution in Canada” (*Uneasy Juncture* 878). The perspective of Battista et al highlights the rhetorical task inherent in writing guidelines. The authors illuminate the challenge of drafting guidelines that base recommendations on evidence but stop short of stripping away the art of physician judgment informed by physician-patient interaction (*Between* 385).

Ultimately the greatest impact on patient care comes from a guideline’s recommendations, so the wording of recommendations is arguably the most important part of the drafters’ task, and

it may be the most difficult. It is certainly the most rhetorical guideline draft task, as the meaning of recommendations emerges from the person reading it. The CDC identifies 3 audiences for its guidelines: patients, clinicians, and policymakers, and sets explicit communication goals for each:

For patients, guideline categories convey desirability of the course of action. Category I means “Most people in the patient’s situation would want the recommended course of action and only a small proportion would not.” Category II means “Most people in the patient’s situation would want the recommended course of action, but some may not” (US CDC 17).

To clinicians, categories convey level of clinical imperative for a course of action: Category I means “Most patients should receive the recommended course of action.” Category II means “Different choices will be appropriate for different patients and clinicians must help patients arrive at management decisions consistent with her or his values and preferences” (US CDC 17).

For policymakers, categories indicate level of mandate. According to the CDC, Category I recommendations “may be considered for policy in many situations.” Category II recommendations will require “substantial debate” to be translated into policy (US CDC 17)

The strength of a recommendation is generally linked to the quality of the evidence, with high quality evidence leading to strong recommendations. For Category I recommendations, strongly directive verbs, like “do” or “do not” are used; for Category II recommendations, verbs like “consider” are used (US CDC 18).

Guideline drafters explicitly select verbs to convey the strength of each recommendation.

Generally, the expressed intent of ultrasound practice guidelines is to provide flexible guidance. From the preamble to all AIUM practice guidelines:



Practice guidelines of the AIUM are intended to provide the medical ultrasound community with guidelines for the performance and recording of high-quality ultrasound examinations. The guidelines reflect what the AIUM considers the minimum criteria for a complete examination in each area but are not intended to establish a legal standard of care. AIUM-accredited practices are expected to generally follow the guidelines with recognition that deviations from these guidelines will be needed in some cases, depending on patient needs and available equipment. Practices are encouraged to go beyond the guidelines to provide additional service and information as needed.

But although the overall tone of ultrasound practice guidelines implies flexibility, these are complex documents that, in fact, contain language from all categories, or levels, of recommendations. From the AIUM Practice Guideline for the Performance of the Musculoskeletal Ultrasound Examination, IV. Supervision and Interpretation of Ultrasound Examinations:

“A physician must be available for consultation with the sonographer on a case-by-case basis” (‘standard’ language).

“Ideally, the physician should be on-site and available to participate actively in the ultrasound examination when required. It is recognized, however, that geographic realities may not permit the presence of an on-site physician in all locations.” (‘guideline’ language).

“Throughout the examination of the rotator cuff, the cuff should be compressed to detect nonretracted tears” (‘guideline’ language).

“In the evaluation of rotator cuff tears, comparison with the contralateral side may be useful” (‘option’ language).

If guidelines are to help clinicians and patients make informed decisions, they must meet the basic standards of technical documents: Considering the audience, guideline drafters must

present well-informed, unbiased recommendations in using clear, unambiguous language, using a user-friendly format.

But other factors affect recommendation strength as well. For example, during the Avian flu epidemic, as soon as an effective treatment was discovered, a strong recommendation was issued to provide that treatment to patients. No randomized controlled trial data with high quality data supported the recommendation, but the risk of fatality and of spread of the disease was great. Delaying treatment until it could be proven effective was not practical, and randomized control trials would have been unconscionable.

### **Emerging Users and Appropriateness**

When new approaches enter the patient care arena, they have not yet generated much evidence from which to develop recommendations, and guideline development for evolving ultrasound applications is controversial. The evidence problem associated with writing CPGs for emerging applications was addressed by the 1993 Canadian Workshop on Clinical Practice Guidelines. A summary of this group's proceedings reports workshop participants' discussions of challenges and best approaches to developing clinical practice guidelines. One issue raised at the workshop was the problem of writing guideline statements for newly evolving practices. Although some members of the group did not support development of guidelines when evidence was sparse, others felt that "scientific uncertainty" should not impede CPG development. Attendees concluded that even when little relevant evidence exists and definitive statements are not possible, guidelines can be written based on expert opinion, provided that the quality of supporting evidence is properly characterized. To enhance validity of such guidelines, in the

absence of substantial evidence these authors recommend a thorough explanation of methods, with revisions as necessary to develop growing knowledge (1715-1719).

Most importantly, the document must provide evidence that every reasonable effort was made to overcome bias. When recommendations proceed largely or entirely on expert opinion, validity must be established through these features:

- Guidelines development committee composition to include a broad spectrum of perspectives
- A well-documented, high quality, transparent collaborative process
- Accurate characterization of evidence and recommendations
- A built in method of gathering evidence prospectively to validate recommendations or to inform revisions

Summarize the lit review, and lead into

.....  
.....



## **CHAPTER THREE: METHODOLOGY**

In this chapter I present the criteria I used to analyze and rate technical writing attributes of several current CPGs. The quality of these attributes influences the accuracy and readability of guideline documents, and so affects their respective values as clinical tools. For a CPG to be valuable to clinicians, and ultimately to patients, it must convey accurately characterized information in clear, unambiguous language in a format and style that makes the information accessible to the user. I propose that when a guideline meets the criteria I set forth in this chapter, the guideline writing quality promotes the users' needs for validity and usability. it 1) meets the definition of validity and 2) it meets usability standards.

To begin, I discuss the background and the writing-specific features of Appraisal of Guidelines Research in Evaluation (AGREE) instrument. Because of its demonstrated value and its suitability to assessing technical communication, AGREE served as the foundation for the rubric I developed. Next I describe variations in purpose and functionality among guidelines, and I examine a guideline's functional "category" as a crucial factor in the quality equation. Finally, I advance a guideline assessment rubric I developed by combining selected items from AGREE with additional criteria related to the nature of each category. In the chapter that follows, I use this new tool to evaluate the quality of several CPGs.

### **The AGREE instrument**

In the mid 1990s, the EU Biomedicine and Health Research Programme sponsored an international meeting of researchers to consider the problem of variation in the quality of CPGs. Included among the participants were representatives from Canada, Denmark, England, Finland,

France, Germany, Italy, Netherlands, New Zealand, Spain, Scotland, Switzerland, and the United States. These researchers collaboratively developed an instrument designed for use both as a tool for assessing the quality of existing CPGs and as a guide to a “harmonious” development process. Called the Appraisal of Guidelines Research an Evaluation (AGREE), the instrument evaluates the quality of reporting found in CPG documents, and it assesses some aspects of the guideline development process. The AGREE tool was extensively field-tested and validated through application to a random sample of guidelines. AGREE is now looked to as a reference by guideline developers all over the world.

AGREE provides structured assessments for six guideline “quality domains” identified by its developers as fundamental to guideline quality. The quality domains include guidelines’ 1) Scope and purpose, 2) Stakeholder involvement, 3) Rigour of development, 4) Clarity and presentation, 5) Application, and 6) Editorial independence. Within each quality domain are items that contribute to the domain’s objective. A total of 23 items comprise the AGREE instrument.

### **Domain 1: Scope and Purpose**

The first quality domain is that of “Scope and Purpose.” According to the criteria outlined by AGREE, to justify the resources invested in development of a particular guideline, the need for the guideline should be conveyed in its introduction. Also in the introduction, the overall objectives of the guideline, the core clinical questions it will address, and the guideline’s target patients should be stated. It is crucial for these introductory details of scope and purpose to be clearly and unambiguously defined because they will direct the guideline development

research and focus guideline developers' recommendations. To meet AGREE's scope and purpose criteria:

Item 1. Statements should provide a rationale for the guideline, and address its expected impact, i.e., improvement of a particular problem. Content should identify the guideline's expected result, target population, and category of care.

Item 2. The fundamental clinical questions, or interrogative statements, should be detailed, and they should provide operational links to the key recommendations. Content should define the intervention of interest, and when more than one intervention is considered, the options should be compared. The guideline should also describe for whom the interventions are being considered, the expected outcomes, and the treatment setting.

Item 3. Relevant characteristics of the target population should be reported, and characteristics that would exclude some patients should be identified.

## **Domain 2: Stakeholder Involvement**

AGREE's second quality domain, "Stakeholder involvement," assesses how effectively a guideline development group represents the perspectives of patients and other individuals and groups whose interests will be directly affected by the guideline. Underlying this domain's concern is dominance of special interests, and the potential for a biased development process that could undermine the ethics and practicality of guideline application. For example, fetal echocardiographic examinations are performed and interpreted by members of several different clinical groups, including pediatric cardiologists, maternal fetal medicine specialists, and specially trained obstetricians. If a guideline for performing fetal echoes is developed by a committee that does not include members of one of those professional groups, the guideline may

be weakened by omission of the group's particular training pathway or its perspective on examination components. If the insurer's perspective is not considered, costs associated with guideline recommendations could become contentious. If patients' preferences are not considered, there may be poor compliance with recommendations. To meet AGREE's stakeholder involvement criteria:

Item 6. Target user groups should be well defined to include all potentially qualified providers, and the purpose of the guideline should be identified.

### **Domain 3: Rigor of Development**

Quality Domain 3, "Rigor of Development," evaluates a guideline's scientific soundness. This domain deals not with the intrinsic quality of the evidence, but, rather, with the methods through which it is selected, interpreted, and translated into recommendations. The risk of bias and of perpetuating outdated practices is reduced when guideline developers methodically search the literature for relevant evidence sources, and then narrow the number of sources using objective criteria, finally formulating recommendations based on an analysis of the evidence. Thorough, accurate reporting of the development process is fundamental to establishing guideline validity, and regular updating is essential to keeping recommendations aligned with current evidence. To meet AGREE's criteria for developmental rigor:

Item 8. Explicit definition of inclusion/exclusion criteria for evidence should be provided, along with clear statement of rationale for the criteria.

Item 9. Strengths and weaknesses of each piece, or group of pieces, of evidence should be described. Descriptions should focus on its inherent risk of bias, and on the method used by guideline developers to interpret the evidence.



Item 10. The collaborative method used to formulate recommendations should be described, along with outcomes, areas of disagreement, and the level of consensus.

Item 12. Recommendations should be explicitly linked to their supporting evidence. When evidence is sparse and a recommendation is based on expert consensus, the nature of the link should be explicitly stated.

Item 14. A timetable and methodology for guideline update should be provided.

#### **Domain 4: Clarity of Presentation**

The next domain, Clarity of Presentation, evaluates document usability. While background information and evidence analysis are crucial to guideline quality, for CPGs to be useful to busy clinicians these documents must provide answers to their questions, and those answers must be clear, unambiguous, and easy to find. Recommendations should be clear and specific, and all options should be enumerated. When the guideline is lengthy or complex, the most important recommendations should be grouped for easy access. To satisfy the criteria of Domain 4:

Item 15. Recommendations should clearly specify and describe the option which was determined from the evidence to be most appropriate for an identified population with particular indications. Any uncertainty about recommendations should be disclosed in the guideline.

Item 16. Management guidelines should present all possible options and provide guidance as to which option is most appropriate for each condition variation.

Item 17. Key recommendations should answer the central clinical questions. These answers should be easy to find in the document.

## **Domain 5: Applicability**

The 5<sup>th</sup> domain of Applicability measures guideline features that promote its use and usefulness. Guideline criteria may help clinicians assess and improve the quality of their care practices, and they may be useful as measures of guideline effectiveness.

Item 19. To improve guideline effectiveness, advice and/or tools to enhance dissemination and implementation should be provided

Item 21. To make guidelines part of an iterative health improvement process they should include criteria and tools for measuring their use and impact. Effective criteria should proceed directly from the key recommendations.

## **Domain 6: Editorial Independence**

An issue in the final domain, Editorial Independence, is the concern for commercial influences that may bias guideline developers. Interaction between clinicians and medical industries is essential to advancing medical technology and pharmacology, but close commercial relationships may color clinicians' preferences when considering options. To comply with AGREE's Editorial Independence criteria:

Item 23. Members of a development group must declare any personal connection that could bias their interpretation of the evidence. Any such declarations should be explicitly stated.

## **Rubric Development**

### **Rubric Version 1 (RV1)**

As a first step in developing my rubric, I adopted 15 attributes from AGREE (Figure 1). Items selected for inclusion in this first version of my rubric focused on document composition; items focused entirely on methodology were omitted. For each guideline evaluated, the included attributes will be rated for writing effectiveness on a 4-point scale, from “Strongly Disagree (1)” to “Strongly Agree (4),” with comments as indicated.

RUBRIC VERSION 1					
ATTRIBUTE	1	2	3	4	COMMENTS
Domain 1. Scope and Purpose					
1. Guideline objectives					
2. Clinical question(s)					
3. Target population					
4. <i>AGREE item omitted</i>					
Domain 2. Stakeholder Involvement					
5. <i>AGREE item omitted</i>					
6. Target users					
Domain 3 Rigor of Development					
7. <i>AGREE item omitted</i>					
8. Evidence selection criteria					
9. Evidence strengths, limitations					
10. Method for formulating recommendations					
11. <i>AGREE item omitted</i>					
12. Link between evidence & recommendations					
13. <i>AGREE item omitted</i>					
14. Updating procedure					
Domain 4. Clarity of Presentation					
15. Recommendations					
16. Management options					
17. Key recommendations					
Domain 5. Applicability					
18. <i>AGREE item omitted</i>					
19. Implementation tools					
20. <i>AGREE item omitted</i>					
21. Monitoring &/or audit criteria					
Domain 6. Editorial Independence					
22. <i>AGREE item omitted</i>					
23. Competing interests of CPG development group					

Figure 1: Rubric Version 1 (RV1)

### **Guideline categories: Qualifying the rubric**

NGC currently publishes on its website summaries of more than 2500 CPGs. At their cores, CPGs all serve one overarching purpose: to make valid recommendations that will assist practitioners with health care decision-making. But CPGs address a broad spectrum of services whose basic purposes and complexities differ, and I propose that these differences need to be considered in the construction of guideline assessment tools.

Based upon the primary tasks addressed by each guideline, the NGC defines eleven guideline categories. NGC Guideline Categories include:

**Assessment of Therapeutic Effectiveness.** This guideline's function is to measure the benefit of the practice in question, considering the effectiveness of the practice and its acceptability to the patient.

**Counseling.** This guideline's function is to recommend methods professionals may use to educate or guide patients to better physical or emotional health.

**Diagnosis.** The function of this guideline is to recommend ways to identify and differentiate clinical conditions. Guidelines in this category may describe a multi-step process to include one or more branching decision junctures.

**Evaluation.** An evaluation guideline outlines a process for gathering information from which a diagnosis and a possible treatment plan may be developed, or later in treatment to assess its effectiveness.

**Management.** Management guidelines describe a comprehensive process of integrating and implementing treatment and evaluation.

Prevention. These guidelines recommend means to promote health and prevent physical and mental disorders

Rehabilitation. These guidelines recommend interventions and techniques for restoring well-being.

Risk Assessment. These guidelines recommend ways to assess the potential of harm resulting from a negative health factor or the lack of a beneficial factor.

Screening. Guidelines in this category recommend means of early detection of disease or the increased likelihood of disease.

Technology Assessment . These guidelines evaluate the use of a particular technology for a clinical purpose, comparing it to other technologies used for the same application.

Treatment. These guidelines recommend interventions meant to relieve illness or injury.

Although the NGC describes these eleven separate “Guideline Categories,” in fact, most guidelines are comprised of elements of more than one category, with most incorporating “Treatment” (2481), “Management” (2240), and/or “Evaluation” (2168), along with elements of other categories.

### **Guideline Category Assessment**

I propose that functional differences among these various categories create differences in the types of tasks they address, and that these differences are important variables to consider when judging guideline quality.

To simplify guideline characterization by function, I suggest combining the eleven guideline categories differentiated by the NGC into three larger groups according to general task structures.

## **Category Group 1: Stratified Recommendations**

Of the eleven categories, five comprise guidelines that describe and compare options, stratifying the options by appropriateness:

- Assessment of therapeutic effectiveness
- Diagnosis
- Risk assessment
- Technology assessment
- Treatment

Guidelines in Category Group 1 (CG1) present evidence of relative effectiveness, benefits, and risks of each available option. Based upon strength of evidence, guidelines in this category stratify options and make corresponding practice recommendations. CG1 guidelines essentially make decisions for clinicians, and so should provide compelling evidence that the recommendations they advance are sound. Guidelines in this category may include evidence tables and may define rubrics used to grade evidence and recommendations.

## **Category Group 2: Multi-option Recommendations**

The guidelines in these next 4 categories offer multiple options, not strictly stratified and not mutually exclusive:

- Rehabilitation
- Screening
- Prevention
- Counseling

Guidelines in Category Group 2 (CG2) list options that can be used together, or in some combination based on resources and patients' preferences, to improve or promote well-being. Compared to CG1 guidelines, these guidelines leave more decision-making to the clinician. To support informed clinical decisions, CG2 guidelines should provide evidence of some measure of effectiveness, if not clear superiority, of each listed option for identified clinical variants.

### **Category Group 3: Performance Recommendations**

Guidelines in the last two categories recommend a process for technical performance of a single, or of a bundled, intervention.

- Evaluation
- Management

Guidelines in Category Group 3 (CG3) serve as performance manuals. These guidelines instruct clinicians on how to perform a particular task, or tasks. While there is an important role for evidence in these guidelines, particularly with respect to exam appropriateness and the composite of steps required to complete these tasks, the guideline body – the technical instruction – does not lend itself to “stratified evidence.” Category 3 Guidelines must be based on the experience and opinions of individuals who are experts at these tasks.

In summary, these criteria that relate to guideline categories will be addressed in my rubric:

CG1 Stratified Recommendation guidelines promote one option over others. Stratification should be based on high quality evidence demonstrating clear advantage(s) of the recommended option. The evidence should be presented in clear, unambiguous, unbiased language, with devices that simplify data interpretation.



CG1 guidelines appropriate decision-making, and so should prove superiority of recommended options.

CG2 Multi-Option Recommendation guidelines provide multiple, largely non-stratified, options. Guidelines in this category should facilitate clinicians' decision-making by providing thorough, broadly comprehensive information about each option and its possible application.

CG2 guidelines inform decision-making and so should demonstrate effectiveness and application of recommended options.

CG3 Performance guidelines lead clinicians through processes. These guidelines should conform to the standards of high-quality user manuals.

CG3 guidelines recommend performance methods, and so should establish through references that they are informed by expert opinion, and they should meet the standards of performance manuals.

### **Adjusting the Rubric for Category Differences**

Next I re-evaluated my rubric, considering the characteristics of each Category Group. Weighing the fundamental information and burden of evidence necessary for each, I determined which RV1 attributes related to those guideline requirements. The following table (Table 1) presents my assessment of the relevance of attributes to guideline categories and, in turn, to the measure of their quality.

Table 1: Category Relevance

Domain.Item	Attribute	Category Relevance		
		CG1	CG2	CG3
1.1	Guideline objectives	√	√	√
1.2	Clinical question(s)	√	√	√
1.3	Target population	√	√	√
2.6	Target users	√	√	√
3.8	Evidence selection criteria	√	√	√
3.9	Evidence strengths, limitations	√	√	
3.10	Method for formulating recommendations	√	√	
3.12	Link between evidence & recommendations	√	√	
3.14	Updating procedure	√	√	√
4.15	Recommendations	√	√	√
4.16	Management options	√	√	√
4.17	Key recommendations	√	√	√
5.19	Implementation tools	√	√	√
5.21	Monitoring &/or audit criteria	√	√	√
6.23	Competing interests of CPG development group	√	√	√

According to this assessment, 11 of the RV1 items are relevant to all guideline categories. The remaining items, numbers 3.8, 3.9, 3.10, and 3.12, are necessary for CG1 guidelines to establish superiority of the recommended options, and to CG2 guidelines to demonstrate the effectiveness of recommended options. I propose that those items are not, however, relevant to CG3 guidelines whose function is not to promote options, but, rather, to recommend performance methods.

While CG3 guidelines do not need to meet all of the criteria set forth by AGREE, I propose that, related to their function as performance manuals, guidelines in this category need to meet some requirements not included in AGREE. According to Gerson (362), user manuals should present clearly defined, well-developed steps and exhibit effective document design. These two attributes, essentially an extension of AGREE Domain 4, should be considered when assessing CG3 Guideline quality, so I will add them to my Rubric Version 1. Guidelines in CG3

also need to establish ethos by referencing expert opinion to support recommended procedures, so I will add an item for that attribute as well.

In addition to the three just-described attributes, I propose that one more is essential to the quality of guideline documents in the CG3 group: justification for performance. Particularly for emerging ultrasound applications where evidence of procedure benefit is sparse, establishing appropriateness will be of critical importance to prevent their misuse and overuse.

Finally, after review of several guidelines, I conclude that guideline documents vary in the quality of their organization, which affects their usability. Including the 4 additional quality attributes to be measured in CG3 guidelines, and adding logical presentation as an assessment item for all groups, I propose this revised list of quality attributes. (See Table 2.) Check marks indicate relevant attributes for guidelines in each category.

Table 2: Category-Adjusted Guideline Quality Assessment

Item #	Attribute	Guideline Category		
		CG1	CG2	CG3
1.	Guideline objectives	√	√	√
2.	Health question(s)	√	√	√
3.	Target population	√	√	√
4.	Target users	√	√	√
5.	Evidence selection criteria	√	√	
6.	Evidence strengths, limitations	√	√	
7.	Method for formulating recommendations	√	√	
8.	Link between evidence & recommendations	√	√	
9.	Updating procedure	√	√	√
10.	Recommendations	√	√	√
11.	Management options	√	√	√
12.	Key recommendations	√	√	√
13.	Implementation tools	√	√	√
14.	Monitoring, audit criteria, and future research	√	√	√
15.	Competing interests of CPG development group	√	√	√
16.	Logical presentation	√	√	√
17.	Clearly defined steps			√
18.	Effective document design			√
19.	Expert reference			√
20.	Appropriateness criteria			√

From the assessment rubric just described, I created the following worksheet (See Figure 2) to use when evaluating guidelines in Chapter 4.

Guideline Title:						
Guideline Developer:						
Guideline Category(s): GC1		GC2		GC3		
<i>Instructions:</i>						
Rate Items 1 – 12 for ALL Guidelines. 4 = Strongly Effective 3 = Moderately Effective 2 = Moderately Ineffective 1 = Strongly Ineffective, or Missing GC1 and GC2 Guidelines: Rate Items 13 – 16 * GC3 Guidelines : Rate Items 17 – 20 **						
#	ATTRIBUTE	RATING				COMMENTS
		1	2	3	4	
1	Guideline objectives					
2	Health question(s)					
3	Target population					
4	Target users					
5	Updating procedure					
6	Recommendations					
7	Management options					
8	Key recommendations					
9	Implementation tools					
10	Monitoring, audit criteria, and future research					
11	Competing interests of CPG development group					
12	Logical presentation					
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria					
14	Evidence strengths, limitations					
15	Method for formulating recommendations					
16	Link between evidence & recommendations					
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria					
18	Clearly defined steps					
19	Effective document design					
20	Expert reference					
<i>TOTAL SCORES:</i>						
= _____						

Figure 2: Category-Adjusted Guideline Assessment Tool

In Chapter Four, I present my analyses of six clinical practice guidelines that address various ultrasound applications. Using the rubric above, I assess text components of these documents and rate how effectively each one promotes guideline validity and usability.

## **CHAPTER FOUR: ANALYSIS**

In this chapter I present my analyses of six clinical practice guidelines that address various ultrasound applications. Using the rubric I described in Chapter Three, I assess text components of these documents and rate how effectively the writing supports each attribute and promotes guideline validity and usability. Although I assign a score to attributes and to each document overall, the primary purpose of my assessments is to identify writing practices that weaken guidelines and to highlight practices that make the documents strong. With this study I evaluate features and writing challenges fundamental to CPGs, but I direct special focus toward those attributes that may prove most important, and most challenging, for developers of guidelines for emerging ultrasound applications. I believe findings and conclusions from these analyses may be used by guideline drafters to enhance the quality of CPGs for emerging users.

### **Guideline Assessments**

The six CPGs I selected all address topics related to ultrasound, but they span a broad spectrum of settings, target user groups, and types of applications. Where two guidelines are published by the same organization, I review them together, describing common features and noting differences.

#### **Guideline 1: American College of Emergency Physicians**

*Critical issues in the initial evaluation and management of patients presenting to the emergency department in early pregnancy (2003)*

The first two guidelines I review are published by the American College of Emergency Physicians (ACEP). The first of these, the earliest published among this group, is ACEP's

Clinical Policy: Critical Issues in the Initial Evaluation and Management of Patients Presenting to the Emergency Department in Early Pregnancy. This ten-page guideline written in 2003 describes point-of-care ultrasound as one component of a diagnostic plan. The guideline is composed of seven sections. In the first section, the “Introduction,” the guideline’s rationale and critical issues are provided. Section 2, “Methodology,” describes the literature review method and criteria for ranking evidence and for grading recommendations. Section 3, “Critical Questions,” presents diagnostic questions related to interpretation of blood pregnancy test results. For each question, recommendations follow a review of relevant evidence. Two of these three questions relate blood test results to findings from point-of-care ultrasound examination. The 4<sup>th</sup> and 5<sup>th</sup> sections, “Methotrexate in Ectopic Pregnancy” and “Rh Seroconversion and Indications for Anti-D Immunoglobulin,” appropriateness of these treatments is questioned. Following a review of the evidence for each, patient management recommendations are provided. No questions are identified for sections 6, “Threatened or Complete Abortion or Ectopic Pregnancy,” or 7, “Minor Abdominal Trauma,” but evidence related to the conditions is discussed, with recommendations following. After the body of the guideline, members of the document development group are listed. References listed at the end of the guideline are also embedded in the text to indicate support for evidence statements.

Using the Category-Adjusted Guideline Quality Assessment Worksheet I prepared in Chapter 3, I rated attributes of this guideline. The findings are presented below in Table 3.

Table 3: Clinical Policy: Critical Issues in the Initial Evaluation and Management of Patients Presenting to the Emergency Department in Early Pregnancy (2003)

<i>Guideline Developer:</i> American College of Emergency Physicians						
<i>Guideline Category(s):</i> GC1 <u>  x  </u> GC2 _____ GC3 _____						
<i>Instructions:</i>						
Rate Items 1 – 12 for ALL Guidelines. 4 = Strongly Effective 3 = Moderately Effective 2 = Moderately Ineffective 1 = Strongly Ineffective, or Missing GC1 and GC2 Guidelines: Rate Items 13 – 16 * GC3 Guidelines : Rate Items 17 – 20 **						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives	x				Clear statement of clinical imperative and topics. (see comments below)
2	Health question(s)	x				“Critical Questions” are explicit, bolded, prominently located, and numbered for easy identification. (see comments below)
3	Target population	x				The title specifies the target population.
4	Target users	x				The target user group is defined (see comments below)
5	Updating procedure	x				A plan for guideline review and updating was provided. (see comments below)
6	Recommendations	x				Recommendations use unambiguous, directive language and directly state justification for each recommendation. Strength grades are included with each recommendation. (see comments below)
7	Management options	x				For a particular setting, 2 options were discussed, and option selection guidance was provided. (see comments below)



#	ATTRIBUTE	RATING				COMMENTS
8	Key recommendations	x				Key recommendations appear as responses to specific health questions. Recommendations are easily found under informative, bolded headings.
9	Implementation tools			x		The recommendations are easy to access, but the guideline could be enhanced with additional implementation tools.
10	Monitoring, audit criteria, and future research				x	None of the recommendations meet the developers' criteria for Level A strength, so the value of the guideline might be enhanced by specific suggestions for further research.
11	Competing interests of CPG development group				x	No conflict of interest disclosures appear in the guideline.
12	Logical presentation	x				Easy to follow progression from clinical question through evidence to recommendations.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria	x				Details of search and selection process are provided. (see comments below)
14	Evidence strengths, limitations	x				The method of interpreting evidence and guarding against biased recommendations was described. (see comments below)
15	Method for formulating recommendations	x				The method is explicitly described. (see comments below)
16	Link between evidence & recommendations	x				Each clinical question is explicitly answered with review of relevant literature, and recommendations are provided as conclusions from the evidence. (see comments below)
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria					
18	Clearly defined steps					
19	Effective document design					
20	Expert reference					
<b>TOTAL SCORES:</b>		5	2	1		= _____ 55 / 64 _____
		2				

## COMMENTS

### 1. *Guideline objectives.*

“In the past decade, despite major changes in epidemiology, incidence, and demographic, ectopic pregnancy remains the most common cause of maternal death and serious morbidity in the first trimester of pregnancy. For this reason, critical issues selected for this policy revision are those primarily associated with initial evaluation and management of ectopic pregnancy... Those critical issues include: (1) ... , (2)... , and (3) ... .”

“This policy revision presents evidence for answering important questions about these critical diagnostic and management issues.”

### 2. *Clinical questions.*

“Is transvaginal ultrasound useful in detecting intrauterine pregnancy when the serum hCG level is less than 1,000 mIU/mL?”

“Is transvaginal ultrasound useful in detecting ectopic pregnancy when the serum hCG level is less than 1,000 mIU/mL?”

“What is the role...?” etc.

### 4. *Target users.*

“This guideline is intended for physicians working in hospital-based emergency departments.”

### 5. *Updating procedure.*

“Clinical policies are scheduled for revision every 3 years; however, interim reviews are conducted when technology or the practice environment changes significantly.”

6. *Recommendations.*

“Level B recommendations. Obtain a repeat serum hcG determination at least 2 days after the initial presentation because it is useful in characterizing the risk of ectopic pregnancy and the probability of a viable intrauterine pregnancy.”

7. *Management options.*

“Because the symptoms associated with gastrointestinal side effects of methotrexate therapy may mimic an acute ectopic rupture, rule out ectopic rupture resulting from treatment failure before attributing gastrointestinal symptoms to methotrexate toxicity. “

13. *Evidence selection criteria.*

“Abstracts and articles were reviewed by subcommittee members, and pertinent articles were selected. These were evaluated, and articles addressing the questions considered in this document were chosen ...”

14. *Evidence strengths/limitations.*

“All publications were graded by at least 2 of the subcommittee members...”

“Factors such as heterogeneity of results, uncertainty about effect magnitude and consequences, strength of prior beliefs, and publication bias...might lead to a downgrading of recommendations.”

15. *Method for formulating recommendations.*

“During the review process, all articles were given a baseline ‘strength of evidence’ by the subcommittee members according to [explicitly stated] criteria.” “Recommendations regarding patient management were then made according to [explicitly stated] criteria.”

16. *Link between evidence and recommendations.*

The clinical question: “Is transvaginal ultrasound useful in detecting intrauterine pregnancy when the serum hCG level is less than 1,000 mIU/mL?”

The evidence: From cited study #1: “All 9 pregnancies were correctly identified ... from 300 mIU/mL and 1,000 mIU/mL” From study #2: “... correctly diagnosed ... 800 mIU/mL.” From additional studies similar evidence was reported.

The recommendation: “Consider transvaginal ultrasound because it may detect intrauterine pregnancy when the serum hCG level is below 1,000 mIU/mL.”

## **Guideline 2. American College of Emergency Physicians**

*Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients With Suspected Appendicitis (2010)*

The second ACEP guideline, Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department (ED) Patients With Suspected Appendicitis was published 7 years later in 2010. In this much lengthier 35-page document the authors describe a management plan for patients who visit the ED with complaints of acute abdominal pain. In the “Introduction” and “Methodology” sections of this guideline, the information provided is similar in type to the information provided in corresponding sections of the early pregnancy guideline.

There are, however, some important differences between the two guidelines. ACEP's 2002 guideline on managing pregnant patients in the ER lists the names of development group members at the end of the guideline; this 2010 guideline from ACEP lists the development group on the cover, with their titles expanded to include their titles and institutional affiliations. Among

the 2002 developers, no one is described as a methodologist; from the 2010 group, two methodologists are identified.

The format for reporting recommendations has evolved as well. In the 2002 guideline each explicit health question is followed by a review of the literature that addresses the question. Resulting recommendations follow the literature review. The format of the 2010 guideline enhances usability by providing the recommendations immediately after the questions they answer, with the literature review following the recommendations. While the format used by guideline developers in 2002 was logical, the newer format is more user-friendly. In addition to the narrative lit reviews in the body of the guideline, the 2010 guideline provides study analyses in evidentiary tables. A comparison of these two guidelines reveals maturation of the guideline development process with attention to easier user access and greater transparency.

In Table 4, below, I present attribute ratings for the second guideline.

Table 4: Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients With Suspected Appendicitis (2010)

<i>Guideline Developer:</i> American College of Emergency Physicians						
<i>Guideline Category(s):</i> GC1 <u>  x  </u> GC2 <u>  X  </u> GC3 <u>      </u>						
<p><i>Instructions:</i>            Rate Items 1 – 12 for ALL Guidelines.            4 = Strongly Effective            3 = Moderately Effective            2 = Moderately Ineffective            1 = Strongly Ineffective, or Missing            GC1 and GC2 Guidelines: Rate Items 13 – 16 *            GC3 Guidelines : Rate Items 17 – 20 **</p>						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives	x				The diagnostic challenge, and clinical and legal impacts of the clinical problem are discussed; Guideline goals are stated. (see comments below)
2	Health question(s)	x				Questions define patient groups and imaging options considered, and they provide framework for comparisons. Questions are prominently located, bolded, and numbered for easy identification. (see comments below)
3	Target population	x				Along with the overall target population, subgroups of interest are identified, as well as excluded groups. (see comments below)
4	Target users	x				The intended user group is defined, and all groups who contributed to development of the CPG were listed. Intended use of the guideline is described. (see comments below)
5	Updating procedure	x				A plan for guideline review and updating was provided. (see comments below)

#	ATTRIBUTE	RATING			COMMENTS
6	Recommendations	x			Recommendations are graded for strength; justification is provided in the literature review immediately following. Recommendations are written using unambiguous, directive language. (see comments below)
7	Management options	x			For a particular setting, 2 options were discussed and option selection guidance was provided. (see comments below)
8	Key recommendations	x			Key recommendations immediately follow questions to which they respond. Recommendations are easily found under informative, bolded headings.
9	Implementation tools			x	The recommendations are easy to access, but the guideline could be enhanced with additional implementation tools.
10	Monitoring, audit criteria, and future research	x			Future areas of research are suggested to improve answers to the guideline's questions. (see comments below)
11	Competing interests of CPG development group	x			The guideline contains a bolded conflict of interest disclosure. (see comments below)
12	Logical presentation	x			Direct, uncluttered connection between questions and recommendations.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>					
13	Evidence selection criteria	x			The Methodology section details the literature search and selection parameters.
14	Evidence strengths, limitations	x			Evidence is reviewed both as a narrative to answer each question, and in evidentiary tables. (see comments below)
15	Method for formulating recommendations	x			The method is explicitly described. (see comments below)
16	Link between evidence & recommendations	x			Guideline drafters' interpretive link between evidence and recommendations is reported. (see comments below)

#	ATTRIBUTE	RATING			COMMENTS
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>					
17	Appropriateness criteria				
18	Clearly defined steps				
19	Effective physical design				
20	Source of expertise identified				
<i>TOTAL SCORES:</i>		60		2	= _____ 62 / 64 _____

### *COMMENTS*

1. *Guideline objectives.*

“Abdominal pain is a high-volume, high-risk chief complaint... The diagnosis of appendicitis can be challenging even in the most experienced of clinical hands.”

“A writing subcommittee reviewed the literature to derive evidence-based recommendations to help clinicians answer [these] critical questions [related to management of emergency department patients with abdominal pain]: (1)..., (2) ..., (3) ... .”

2. *Health questions.*

“Can clinical findings ... guide decision-making...?”

“... what is the role of contrast?”

“... what are the roles of CT and ultrasound in diagnosing acute appendicitis?”

3. *Target population.*

“This guideline is intended for patients presenting to the ED with acute, nontraumatic abdominal pain and possible or suspected appendicitis.”

“In adult patients with suspected acute appendicitis...”

“In children with suspected acute appendicitis...”



“... not intended to address the care of patients with trauma-related abdominal pain, or pregnant patients.”

4. *Target users.*

“This guideline is intended for physicians working in hospital-based EDs.”

“This policy is not intended to be a complete manual on the evaluation and management of patients with nontraumatic acute abdominal pain but rather a focused examination of critical issues that have particular relevance to the current practice of emergency medicine...this guideline defines for the physician those strategies for which medical literature exists to provide support for answers to the crucial questions addressed in this policy.”

5. *Updating procedure.*

“Clinical policies are scheduled for revision every 3 years; however, interim reviews are conducted when technology or the practice environment changes significantly.”

6. *Recommendations.*

“In children, use ultrasound to confirm acute appendicitis but not to definitively exclude acute appendicitis.

7. *Management options.*

“Given the concern over exposing children to ionizing radiation, consider using ultrasound as the initial imaging modality. In cases in which the diagnosis remains uncertain after ultrasound, CT may be performed.”

10. *Monitoring &/or audit criteria/Research.*

Guideline question 2”In adult patients with suspected acute appendicitis who are undergoing a CT scan, what is the role of contrast?”

Research suggestion 1: “A prospective comparison of CT with no contrast, IV contrast alone, and oral and IV contrast (for appendicitis), using the newest CTG technology. This study could be done in both adult and pediatric populations.”

11. *Competing interests of CPG development group.*

“There were no relevant industry relationships disclosed by the subcommittee members...”

14. Evidence strengths, limitations.

In evidentiary tables each study is analyzed by year, design, intervention, outcome measure, results, limitations, and class.

15. Method for formulating recommendations.

“All articles used in the formulation of this clinical policy were graded by at least 2 subcommittee members for strength of evidence and classified by the subcommittee members in to 3 classes of evidence on the basis of...”

“Articles were then grade on 6 dimensions thought to be most relevant to the development of a clinical guideline: ...”

“Articles received a final grade ...”

“Clinical findings and strength of recommendations regarding patient management were then made according to the following criteria:

Level A recommendations...

Level B...

Level C ...”

16. *Link between evidence & recommendations.*

“Theoretically, higher Alvarado scores are associated with a higher likelihood of appendicitis and lower scores with a lower likelihood of appendicitis. Whether the Alvarado score can reliably predict the need for CT is debatable. [cited studies discussed] The authors concluded that patients with scores of 3 or less should not have CT, those with scores between 4 and 6 should have CT, and those with scores 7 or higher... The authors recommended imaging even patients with low Alvarado scores.”

### **Guideline 3: American Urological Association**

*Consensus Statement on Urologic Ultrasound Utilization (2007) (Analysis: Table 5)*

The third guideline was first published by the American Urological Association (AUA) in 1993 and periodically revised, most recently in 2007. The guideline is presented in five sections: Section 1, the “AUA Policy Statement on Imaging Services,” affirms urologists as qualified to perform ultrasound as part of their patient care. “Section 2, “Equipment, Documentation, Indications,” is the operational part of the guideline. This section begins with general recommendations for equipment and for documentation standards, and then follows with targeted details, technique instructions, and indications for six ultrasound exams performed by these users. This largest section is followed by a discussion of “Educational Requirements” set forth by this organization for their members who perform ultrasound, and it details the curriculum of the ultrasound component of Urology Residency Education. Next, Section Five, “Patient Safety,” provides a statement about responsible use of ultrasound.

This guideline is generally easy to follow with bulleted lists and good use of white space.

Its explanations of technique, however, are embedded in text, rather than presented in a more user-friendly list format. Table 5 reports attribute ratings.

Table 5: Consensus Statement on Urologic Ultrasound Utilization (2007)

<i>Guideline Developer:</i> American Urological Association (2007)						
<i>Guideline Category(s):</i> GC1                      GC2                      GC3                      X						
<i>Instructions:</i> Rate Items 1 – 12 for ALL Guidelines. 4 = Strongly Effective 3 = Moderately Effective 2 = Moderately Ineffective 1 = Strongly Ineffective, or Missing GC1 and GC2 Guidelines: Rate Items 13 – 16 * GC3 Guidelines : Rate Items 17 – 20 **						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives			x		Objectives are not clearly defined.
2	Health question(s)			x		No clinical questions are provided to link exam indications to recommended practices
3	Target population			x		The population is characterized per indications, but no other patient descriptions are provided.
4	Target users	x				Urologists are identified as target users; suitability of these users is strongly affirmed. (see comments below)
5	Updating procedure				x	No plan is provided.
6	Recommendations	x				Equipment, performance, and documentation recommendations are provided, and their relative strengths are characterized. (see comments below)
7	Management options	x				When more than one approach may be used, all are clearly stated. (see comments below)

#	ATTRIBUTE	RATING				COMMENTS
8	Key recommendations				x	Recommendations are described, but none are easily identified as key answers to clinical questions.
9	Implementation tools				x	No guideline implementation tools are provided or suggested.
10	Monitoring, audit criteria, and future research				x	No assessment criteria are provided. (see comments below)
11	Competing interests of CPG development group				x	Conflict of interest statements are not provided.
12	Logical presentation	x				Global statements begin and end the guideline, with individual exams grouped under descriptive headings.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria					
14	Evidence strengths, limitations					
15	Method for formulating recommendations					
16	Link between evidence & recommendations					
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria		x			Provides opinion that patient care is optimized when urologists perform their patients' ultrasound exams, but does not offer supporting literature. (see comments below)
18	Clearly defined, ordered steps		x			Steps are clearly stated and concise, but they are grouped in paragraphs, making the beginning of each step difficult to locate.
19	Effective document design			x		Instructions are compressed into dense paragraphs. The user would follow the instructions more easily if steps were separated and numbered. No graphics are included.
20	Expert reference	x				AUA, an organization of urologists, is defined as the guideline developer.
<b>TOTAL SCORES :</b>		2 0	4	8	5	= _____ 37 / 64 _____

## COMMENTS

### 4. *Target users.*

“The AUA affirms that urologists are the physicians best qualified to diagnose, manage and treat diseases and conditions of the genitourinary tract in patients of all ages.”

“Urologists combine the technical skill in the use of imaging equipment with the cognitive skills of the underlying disease process...”

“The acquisition and maintenance of skills and knowledge associated with imaging technology is assured by the Accreditation Council for Graduate Medical Education residency review committee for urology...”

Note: Section 3, Educational Requirements, details the ultrasound training residents receive, listing didactic areas of instruction and describing hands-on training. Post-residency training is also described, and the extent of ultrasound knowledge tested for urology certification is characterized. This method of defining training guidelines may be of value as a model for other emerging user groups.

### 6. *Recommendations*

“The equipment should display mechanical and thermal indices and provide for adjusting power output.”

“An antibiotic may be administered if biopsy is anticipated.”

“A moderately filled bladder is necessary for optimal imaging of the bladder.”

### 7. *Management options*

“The needle guide may be a single-use device or a reusable device.”

“The bladder should be scanned in the transverse and sagittal views. Bladder imaging can also be performed using [approach 1] or [approach 2]. It may also be performed by [approach 3] using special ultrasound transducers...”

10. *Monitoring &/or audit criteria*

Note: Patient satisfaction and outcomes studies could be recommended to evaluate appropriateness of urologists performing ultrasound exams

17. *Appropriateness criteria*

“Urologists integrate an understanding of the risks and benefits of imaging technologies with the clinical care of the patient. Patient care is optimized when urologists coordinate the use of appropriate imaging techniques and equipment in the setting most beneficial to their patients.”

**Guideline 4: American College of Cardiology Foundation, et al**

*2008 Appropriateness Criteria for Stress Echocardiography: A Report of the American College of Cardiology Foundation Appropriateness Criteria Task Force, American Society of Echocardiography, American College of Emergency Physicians, American Heart Association, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the Heart Rhythm Society and the Society of Critical Care Medicine. (Analysis: Table 6)*

The fourth guideline explores appropriateness concerns for this traditional ultrasound application. The text of this 22-page strongly collaborative document is preceded by a list of guideline developers and their affiliations.

This guideline addresses the dual tasks of promoting optimal patient outcomes and maintaining cost-effectiveness. Following an Abstract, the document contains eight major sections, including ten tables. In the first section, the “Preface,” the guideline objective is stated, along with a description of the collaborative appropriateness-rating process employed by the committee. The second section, the “Introduction,” explains the guideline’s rationale, and in the next section, “Methods,” the drafters describe the formal consensus method used to reduce bias in this largely expert opinion-based guideline. The fourth section defines abbreviations. The fifth section provides a general discussion of Results of Ratings, followed in sections six and seven with reports of scores of Stress Echocardiography Appropriateness Criteria by Indication and by Appropriateness Category, respectively. Section Eight provides a General Discussion of the usefulness and limitations of appropriateness criteria, and of potentially valuable research projects. Finally, the guideline concludes with references and four appendices. Appendices include an assessment algorithm and financial disclosures, and expand on definitions, guideline development methods, and compositions of working groups.

Stress echocardiographic examinations, the subject of this guideline, have been performed by “traditional” users for more than a decade, but the major challenge inherent in development of this CPG underlies guideline development for emerging applications: scarce data, compelling a need to rely largely on expert opinion to generate recommendations. Methods used by these guideline drafters to reduce bias and promote validity may be useful to drafters of guidelines for emerging ultrasound applications.

Exam appropriateness is also likely to be a crucial issue for emerging ultrasound applications, and the introduction to this guideline effectively summarizes the issue. According to these authors, technological improvements have increased ultrasound’s diagnostic utility, but



along with that benefit have come increased possibilities of misuse and overuse of the modality. The authors raise concerns that this ultrasound examination may be prescribed for patients who will not benefit from it, or who may benefit but could have been managed effectively without it. Beyond the immediate expense of an unnecessary exam, the authors point out the danger of exam results prompting “harmful and costly downstream testing” or unnecessary follow-up exams. Rating of this guideline’s attributes follow in Table 6.

Table 6: Appropriateness Criteria for Stress Echocardiography (2008)

<p><i>Guideline Developer:</i> American College of Cardiology Foundation Appropriateness Criteria Task Force, American Society of Echocardiography, American College of Emergency Physicians, American Heart Association, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance.</p>						
<p><i>Guideline Category:</i>    GC1    <u>  x  </u>                    GC2                    GC3</p>						
<p><i>Instructions:</i>            Rate Items 1 – 12 for ALL Guidelines.                4 = Strongly Effective                3 = Moderately Effective                2 = Moderately Ineffective                1 = Strongly Ineffective, or Missing            GC1 and GC2 Guidelines: Rate Items 13 – 16 *            GC3 Guidelines : Rate Items 17 – 20 **</p>						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives	x				Objectives and expected outcomes from the guideline are described (See comments below)
2	Health question(s)	x				Questions concerning appropriateness of the procedure for identified indications are provided graphically in tabled lists. (See comments below)
3	Target population	x				The target population is broadly defined, and then differentiated into several smaller groups according to their symptoms and clinical conditions. (See comments below)
4	Target users	x				Clinicians, facilities, and payers are identified as target users, and suggestions are provided as to how the document might be used. (See comments below)
5	Updating procedure			x		Necessity for updating was acknowledged, but no plan was provided. (See comments below)

#	ATTRIBUTE	RATING				COMMENTS
6	Recommendations	x				Recommendations take the form of quantitative appropriateness rating of the procedure for each clinical variant. (See comments below)
7	Management options	x				The document provides guidance on how to use appropriateness scores.
8	Key recommendations	x				Appropriateness ratings are easily found in the guideline's tables.
9	Implementation tools	x				The tables are well-formatted and easy to reference.
10	Monitoring, audit criteria, and future research	x				Areas wherein future research is needed are described. (See comments below)
11	Competing interests of CPG development group	x				A strong, specific statement addresses potential relationships between CPG developers and industrial interests. Additionally, each author's professional affiliation and role in the CPG development process is listed. (See comments below)
12	Logical presentation	x				A table of contents demonstrates a logical, easily accessible document structure.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria	x				"Relevant guidelines," along with their reference lists, were identified as the sources of evidence. Key words are provided with the reference list.
14	Evidence strengths, limitations			x		The considered evidence is qualitatively described as "relevant," "objective," and "nonbiased," but no formal evidence-grading system is provided. (See comments below)
15	Method for formulating recommendations	x				This guideline draws strongly on judgment and expert opinion for its recommendations. The committee's formal use of a modified Delphi exercise is reported, and other steps the committee took to reduce the risk of bias are described. (See comments below)

#	ATTRIBUTE	RATING				COMMENTS
16	Link between evidence & recommendations	x				The role of expert consensus as the link is explicitly stated. (See comments below)
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria					
18	Clearly defined steps					
19	Effective document design					
20	Expert reference					
<i>TOTAL SCORES:</i>		5 6		4		= _____ 60/64 _____

### *COMMENTS*

*1. Guideline objectives.*

“In an effort to respond to the need for the rational use of imaging services in the delivery of high quality care, the American College of Cardiology Foundation (ACCF) has undertaken a process to determine the appropriateness of cardiovascular imaging for selected patient indications.”

“The ultimate objective of appropriateness criteria is to improve patient care and health outcomes in a cost-effective manner but is not intended to ignore ambiguity and nuance intrinsic to clinical decision making.”

“The primary objective of this report is to provide guidance regarding the perceived suitability of stress echocardiography for diverse clinical scenarios.”

*2. Health question(s).*

Comment: Fifty-one indications were evaluated and questioned as to appropriateness. The guideline developers succeeded in presenting this large number of questions concisely by

listing them in tables. The questions were made even more accessible by being grouped according to several variations in the clinical scenario.

3. *Target population.*

“All indications are assumed to apply to adult patients (18 years of age or older).”

Comment: Table headings further differentiate the population.

4. *Target users.*

“Clinicians could use the ratings as a decision support or educational tool when ordering a test or providing a referral to another qualified physician. The criteria also may be used to facilitate discussion with referring clinicians who have patterns of ordering tests for inappropriate indications. Facilities and payers may choose to use the criteria either prospectively in the design of protocols, automated order entry, and pre-authorization procedures, or retrospectively for quality reports. It is hoped that payers will use this document as the basis to inform rational strategies to ensure that their members receive the highest-quality, cost-effective cardiovascular care.”

“When used to assess performance, appropriateness criteria should be applied in conjunction with systems that support quality improvement.”

5. *Updating procedure.*

“It will be necessary to periodically assess and update the indications and criteria as technology evolves and new data and field experience become available.”

6. *Recommendations.*

“Score 7 to 9” – Appropriate

“Score 4 to 6” – Uncertain

“Score 1 to 3” – Inappropriate

7. *Management options.*

“If major risk predictors are present, Figure A1 suggests consideration of coronary angiography and postponing or canceling noncardiac surgery. Once perioperative risk predictors are assessed based on the algorithm, then the surgical risk and patient’s functional stats should be used to establish the need for noninvasive testing.”

10. *Monitoring, audit criteria, and future research.*

“Future research analyzing patient outcomes utilizing indications rated appropriate would help ensure the equitable and efficient allocation of resources for diagnostic studies. Review of medically necessary care may also improve the understanding of regional variations in imaging utilization. Further exploration of the indications rated as “uncertain” will help generate the data required to further define the appropriateness of stress echocardiography. Finally, it will be necessary to periodically assess and update the indications and criteria as technology evolves and new data and field experience become available.”

11. *Competing interests of CPG development group.*

“The ACCF and its partnering organizations rigorously avoid any actual, perceived, or potential conflicts of interest that may arise as a result of an outside relationship or personal interest of a member of the Technical Panel. Specifically, all panelists are asked to provide disclosure statements of all relationships that may be perceived as real or potential conflicts of interest....A table of disclosures of the Technical Panel and Task Force Members can be found in the Appendix D.” (Appendix D presents a table listing for each committee member any research grants, speaking support, stock ownership, board of director affiliations, and industrial consultation relationships.)

Example of author listing: “Pamela S. Douglas, MD, MACC, FAHA, FASE-Lead Author, Appropriateness Criteria for Stress Echocardiography-Past President, ACC, Past President, ASE, Ursula Geller Professor of Research in Cardiovascular Diseases, Duke University Medical Center, Durham, NC.”

14. *Evidence strengths, limitations.*

“Care was taken in providing objective, nonbiased information, including guidelines and key references, to the Technical Panel.”

15. *Method for formulating recommendations.*

“To prevent bias in the scoring process, the Technical Panel deliberately was not comprised solely of specialists in the particular procedure under evaluation. Specialists, while offering important clinical and technical insights, might have a natural tendency to rate the indications within their specialty as more appropriate than nonspecialists. In addition, care was taken in providing objective, nonbiased information, including guidelines and key references, to the Technical Panel.”

“Two rounds of ratings with lively discussion between the ratings did lead to some consensus among panelists. However, further attempts to drive consensus would have diluted true differences in opinion among panelists and, therefore, was not undertaken.”

Comment: Specific areas of disagreement are noted in the discussion of results of ratings.

16. *Link between evidence & recommendations.*

“The indications were constructed by echocardiography experts and modified on the basis of discussions among the Task Force and feedback from independent reviewers and the Technical Panel Whenever possible, indications were mapped to relevant clinical guidelines and key publications/references.”

## **Guideline 5: American Institute of Ultrasound in Medicine**

*Practice Guideline for the Performance of the Musculoskeletal Ultrasound Examination (2007)*

*(Analysis: Table 7)*

The first of two AIUM guidelines I reviewed is the Practice Guideline for the Performance of the Musculoskeletal Ultrasound Examination, published in 2007.

This guideline addresses an ultrasound examination that may be performed in traditional settings by sonographers or physicians, but is commonly performed by physicians as a point-of-care exam. The thirteen-page guideline developed in collaboration with the American College of Radiology, includes eight sections. The first, the “Introduction,” describes the collaborative nature of the guideline and identifies the target users. Standard organization recommendations are provided in Sections Two, Three, Four, Six, and Eight. Section Seven, “Equipment Specification,” describes optimal equipment for this exam.

Section Five, “Specifications for Individual Examinations,” comprises the body of the guideline. In this section, authors recommend examination views and techniques for acquiring those views. Under appropriate subheadings, instructions are provided for examining each structure, and for using ultrasound to guide interventional procedures. At the conclusion of the guideline, the drafters and references are listed. Following in Table 7 I report attribute ratings for this guideline. (EASY TO READ, etc??



Table 7: Practice Guideline for the Performance of the Musculoskeletal Ultrasound Examination (2007)

<i>Guideline Developer:</i> American Institute of Ultrasound in Medicine and the American College of Radiology						
<i>Guideline Category(s)</i> GC1                      GC2                      GC3      x						
<p><i>Instructions:</i></p> <p>Rate Items 1 – 12 for ALL Guidelines.              4 = Strongly Effective              3 = Moderately Effective              2 = Moderately Ineffective              1 = Strongly Ineffective, or Missing</p> <p>GC1 and GC2 Guidelines: Rate Items 13 – 16 *          GC3 Guidelines : Rate Items 17 – 20 **</p>						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives		x			The guideline states a general purpose, but no specific objectives. (see comments below)
2	Health question(s)		x			Health questions are weakly implied throughout the guideline. (see comments below)
3	Target population		x			The target population is indirectly identified according to indications.
4	Target users		x			Physicians and sonographers are mentioned, but users are only identified by the nonspecific term “practitioner.” (see comments below)
5	Updating procedure				x	No plan for updating the document is provided.
6	Recommendations		x			The guideline is highly informative, but most recommendations are weakly stated in nondirective language. (see comments below)
7	Management options	x				Optional positions are suggested to accommodate patient’s conditions; additional views are recommended when conditions warrant. (see comments below)

#	ATTRIBUTE	RATING				COMMENTS
8	Key recommendations			x		Recommendations are embedded within paragraphs, making them difficult to identify. (see comments below)
9	Implementation tools				x	The document does not incorporate any implementation tools.
10	Monitoring, audit criteria, and suggestions for future research				x	None are provided.
11	Competing interests of CPG development group				x	No conflict of interest statement is provided.
12	Logical presentation	x				Examination performance steps are logically ordered.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria					
14	Evidence strengths, limitations					
15	Method for formulating recommendations					
16	Link between evidence & recommendations					
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria				x	No criteria are provided.
18	Clearly presented steps			x		Steps are difficult to pick out where they are embedded in paragraphs, and most are passively worded.
19	Effective document design			x		The overall document is well-organized, but it contains long, dense paragraphs, very little white space, and no graphics.
20	Expert reference		x			The principal drafter was identified. This author is a leading expert in musculoskeletal ultrasound, and including his professional affiliation and/or background would enhance the document's strength.
<b>TOTAL SCORES:</b>		8	18	6	5	= _____ 37 / 64 _____

## COMMENTS

### 1. *Guideline objectives.*

“This guideline has been developed to assist practitioners performing a musculoskeletal ultrasound examination. While it is not possible to detect every abnormality, adherence to the following guidelines will maximize the probability of detecting most abnormalities that occur.”

### 2. *Health question(s).*

“The indications for ultrasound include, but are not limited to, evaluation of shoulder pain or dysfunction.”

“These (previously mentioned) views should be used to determine if the tendon is properly positioned within the bicipital groove, subluxated, dislocated, or torn.”

### 4. *Target users.*

“This guideline has been developed to assist practitioners performing a musculoskeletal ultrasound examination.”

“A physician must be available for consultation with the sonographer on a case-by-case basis. Ideally, the physician should be on-site and available to participate actively in the ultrasound examination when required.”

### 18. *Clearly defined steps.*

“The palm is placed down on the table, or if the patient is supine, the forearm is placed across the abdomen, with the elbow flexed to 90°. The posterior joint space, triceps tendon, olecranon process, and olecranon bursa are assessed.”

## **Guideline 6: American Institute of Ultrasound in Medicine**

*Practice Guideline for the Performance of Fetal Echocardiography (2010) (Analysis: Table 8)*

The second AIUM guideline I reviewed is the *Practice Guideline for the Performance of Fetal Echocardiography*, published collaboratively with the American College of Obstetricians and Gynecologists and the Society for Maternal-Fetal Medicine, three years later in 2010.

Although this is the first AIUM guideline dedicated to this topic, the exam has been performed by sonographers and physicians for several years and fits the criteria of “traditional.”

Many sections of this guideline are similar in purpose to corresponding sections in the Musculoskeletal Guideline, but there are differences in usability features. In contrast to Musculoskeletal’s dense text discussion of recommended views, Fetal Echo lists views in bulleted lists. The Fetal Echo guideline also includes detailed illustrations, referenced in the text. There are also differences in transparency of links between evidence and recommendations. The Fetal Echo guideline contains embedded recommendations, a convention that has only been adopted for AIUM guidelines during the past two years. Also in the Fetal Echo guideline, the authors characterize the major imaging techniques as “Recommended” or “Optional,” depending on other patient conditions. The attribute ratings for this guideline follow in Table 8.

Table 8: Practice Guideline for the Performance of Fetal Echocardiography (2010)

<p><i>Guideline Developer:</i> American Institute of Ultrasound in Medicine, American College of Radiology, American College of Obstetricians and Gynecologists, and the Society for Maternal-Fetal Medicine. Endorsed by the American College of Radiology.</p>						
<p><i>Guideline Category(s):</i> GC1                      GC2                      GC3    x</p>						
<p><i>Instructions:</i>  Rate Items 1 – 12 for ALL Guidelines.  4 = Strongly Effective  3 = Moderately Effective  2 = Moderately Ineffective  1 = Strongly Ineffective, or Missing  GC1 and GC2 Guidelines: Rate Items 13 – 16 *  GC3 Guidelines : Rate Items 17 – 20 **</p>						
#	ATTRIBUTE	RATING				COMMENTS
		4	3	2	1	
1	Guideline objectives		x			Objectives are obliquely stated; the expected outcome of applying the guideline is stated more directly. <i>(see comments below)</i>
2	Health question(s)	x				Conveyed through exam definition and indications <i>(see comments below)</i>
3	Target population	x				The target population is defined as fetuses between 18 and 22 weeks' gestational age, or younger.
4	Target users				x	Target users are not identified.
5	Updating procedure				x	Not provided.
6	Recommendations	x				Recommended sonographic techniques are boldly identified and characterized according to strength. <i>(see comments below)</i>
7	Management options	x				An alternative scan option is suggested to manage a difficult scan; for some fetal conditions, additional exam components are suggested. <i>(see comments below)</i>
8	Key recommendations	x				Scan recommendations are identified clearly as noted in Item 6..
9	Implementation tools	x				Guideline illustrations aid implementation.

#	ATTRIBUTE	RATING				COMMENTS
10	Monitoring, audit criteria, and future research				x	None provided.
11	Competing interests of CPG development group				x	No conflict of interest statements are provided.
12	Logical presentation	x				The guideline is well organized and easy to follow.
<i>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</i>						
13	Evidence selection criteria					
14	Evidence strengths, limitations					
15	Method for formulating recommendations					
16	Link between evidence & recommendations					
<i>**Rate Items 17 – 20 for CG3 guidelines ONLY!</i>						
17	Appropriateness criteria		x			A criteria for judging appropriateness is provided indirectly. (see comments below)
18	Clearly defined steps	x				A “sequential segmental analysis” is provided and structures to include are listed.
19	Effective document design	x				The document is orderly, makes good use of white space, and it includes excellent graphics.
20	Expert reference			x		No information is provided to convey the expertise of the task force members.
<b>TOTAL SCORES:</b>		3 6	6	2	4	= _____48/64_____

### COMMENTS

#### 1. Objectives.

“Accurate prenatal diagnosis offers potential clinical benefit with regard to infant outcomes, especially in those cases that are likely to require prostaglandin infusion to maintain patency of the ductus arteriosus.”

“While it is not possible to detect every abnormality, adherence to the following guideline will maximize the probability of detecting most cases of clinically significant congenital heart disease.”

2. *Health questions*

“Fetal echocardiography is broadly defined as a detailed sonographic evaluation that is used to identify and characterize fetal heart anomalies before delivery.”

6. *Recommendations. (Note: The following are guideline headings)*

Gray Scale Imaging (Recommended)

Doppler Sonography (Optional but Recommended for Suspected Cardiac Flow Abnormalities)

M-Mode Echocardiography (Optional but Recommended for Cardiac Rate or Rhythm Abnormalities.)

Cardiac Biometry (Optional but Can Be Considered in the Presence of Structural Anomalies)

7. *Management options*

“Technical limitations ... can make a detailed heart evaluation very difficult because of acoustic shadowing, especially during the third trimester. It may be necessary to examine the patient at a different time if the heart is poorly visualized.”

“Doppler Sonography (Optional but Recommended for Suspected Cardiac Flow Abnormalities)”

“Cardiac Biometry (Optional but Can Be Considered in the Presence of Structural Anomalies)”

17. *Appropriateness criteria*

“A request for the examination must be originated by a physician or other appropriate health care provider familiar with the patient’s clinical situation and should be consistent with relevant legal and local health care facility requirements.”

### **Summary of Findings**

Following is a review of the scores assigned to documents for each attribute.

4) Guideline objectives:

3/6 were rated “4.” In the other three guidelines, objectives are stated obliquely or not at all. Effective strategies included:

- a clear statement of topics and clinical imperative
- a description of the diagnostic challenge, and the clinical and legal impacts of the clinical problem
- a description of guideline objectives and expected outcomes

5) “Health questions.”

4/6 were rated “4.” Two received lower scores because questions were not provided or were only weakly implied. Characteristics of effectively presented questions included:

- explicit wording,
- bold font,
- numbering,
- prominent location
- links with exam indications and recommended practices.



6) “Target population.”

4/6 were rated “4.” In each of these guidelines, target populations are explicitly defined. In the other two guidelines, they are identified by indications but by no other characteristics.

7) “Target users.”

4/6 were rated “4.” These four guidelines identified the clinical groups considered qualified to perform the exam. In one of these guidelines, groups represented in the development process were identified; another provided suggestions for use of the guideline. In the two guidelines rated below “4,” target users are identified using nonspecific terms, or they are not identified at all.

8) “Updating procedures.”

Only these two guidelines stated a plan or timetable for updating. A third acknowledged the need for updating, but did not provide a plan or timetable.

9) “Recommendations.”

5/6 were rated “4.” The 6<sup>th</sup> guideline is highly informative, but conveys most recommendations in weak, nondirective language. Characteristics of effective presentation included:

- unambiguous, directive language
- strength grading
- justification for each recommendation

10) “Management Options.”

6/6 were rated “4.” All six guidelines recommended optional approaches as warranted.

11) “Key recommendations.”

4/6 were rated “4.” In the other two, the recommendations are not clearly identified as key answers to clinical questions, or the recommendations are embedded within paragraphs, making them difficult to identify. Characteristics of effective approaches included:

- responsiveness to specific health questions
- ease of identification under informative bolded heading
- presentation in tables

12) “Implementation tools.”

2/6 were rated “4.” The other guidelines fail to include or make reference to any implementation tools. Effective techniques included:

- recommendations in well-formatted tables
- illustrations

13) “Monitoring, audit criteria, and future research.”

2/6 were rated “4.” Those two guidelines suggested areas where future research is needed.

14) “Competing interests of CPG development group.”

2/6 were rated “4.” Disclosure strategies included:

- a bolded conflict of interest statement
- a strong, specific statement addressing potential relationships between CPG developers and industrial interests
- listing of each author’s professional affiliation and role in the CPG development process.

15) “Logical presentation.”

6/6 were rated “4.”

Only three of the six guidelines were evaluated for the following attributes:

16) “Evidence selection criteria”

3/3 were rated “4.” In each, the details of the search and selection process are provided.

17) “Evidence strengths, limitations”

2/3 were rated “4.” In the third guideline, the evidence was qualitatively described as “relevant,” “objective,” and “nonbiased,” but no formal evidence-grading system is provided.

Effective presentation strategies included:

- description of the method used to interpret evidence and reduce bias
- review of evidence both as a narrative to answer each question, and in evidentiary tables

18) “Method for formulating recommendations”

3/3 were rated “4.” In each, the method was explicitly described. One of the three drew strongly on judgment and expert opinion for its recommendations. The committee’s formal use of a modified Delphi exercise is reported, and other steps the committee took to reduce the risk of bias are described.

19) “Link between evidence and recommendations”

3/3 were rated “4.” In two, each clinical question is explicitly answered with review of the relevant literature, and recommendations are provided as conclusions from the evidence. In the 3<sup>rd</sup> guideline, the role of expert consensus as the link is explicitly stated.

20) “Appropriateness criteria.”

0/3 guidelines were rated “4”

21) “Clearly presented steps.”

1/3 guidelines were rated “4.” This guideline describes a “sequential segmental analysis” of structures. In the other two guidelines, the steps are clearly stated and concise, but they are passively worded and grouped in paragraphs, making the beginning of each step difficult to locate.

22) “Effective document design”

1/3 was rated “4.” This document displays an orderly format, makes good use of white space, and includes excellent graphics. In the other two documents, instructions are compressed into long, dense paragraphs, with very little white space and no graphics.

23) “Expert Reference”

1/3 was rated “4.” In this document, the expertise of the clinician guideline development group is affirmed. Neither of the others attests to the expertise of the drafters.

## CHAPTER FIVE: CONCLUSIONS

In this concluding chapter I examine the guideline analyses described in Chapter Four and discuss the conclusions I reached regarding the key questions addressed by this thesis:

- How should guidelines for emerging ultrasound applications be assessed?
- Is it appropriate to develop guidelines for these applications?
- If so, how can these guidelines best be constructed for validity and usability?

### **How should guidelines for emerging ultrasound applications be assessed?**

Among the numerous instruments investigated as guideline quality assessment tools, the AGREE instrument has emerged as the most widely accepted. However, the developers of AGREE vaguely advise that for “some” guidelines, certain attributes designated by the instrument as quality components do not apply. As an example, AGREE authors identify “Management options” as a non-universal attribute. “... guidelines narrow in scope may not provide the full range of options for the management of the condition ” (??). In such cases, the authors suggest “having appraisers skip that item in the assessment process or [rate] the item as “1” (absence of information) and [provide] context about the score” (??).

The purpose of my study was to address these vague exceptions in order to more appropriately evaluate guidelines for emerging applications. I theorized that the inconsistencies that limit applicability of AGREE across guidelines could relate to fundamental differences among guideline purposes – differences that could be defined and incorporated into AGREE as adjustment factors to make its rating scheme more universal.

To test my theory, I first considered characteristics that distinguish different types of guidelines. Beginning with the NGC Guideline Category definitions as a primary sorting tool, I grouped NGC Categories into three larger groups according to common characteristics, and considered how the AGREE tool could be adapted to assess different types of guidelines in a more tailored fashion. In particular, I considered the attributes related to evidence since these are problematic for emerging users' guidelines: evidence selection criteria; evidence strengths and limitations; methods for formulating recommendations; and linking evidence with recommendations.

I determined that guidelines that present options should, indeed, provide evidence of validity, and in the case of stratified options, use evidence to substantiate recommendations. Following that theory, guidelines in both CG1 and CG2 should be assessed according to these evidence criteria, and their scores on those attributes should correlate with quality.

For ultrasound performance guidelines, particularly for emerging applications where the science is young and evidence scarce, however, guideline scores on these four AGREE attributes are likely to suggest poor quality. But I contend that “low” scores on these 4 items do not relate to the quality of ultrasound performance guidelines at all. I theorize that convoluting “evidence-based” with “high quality” is misleading. Those 4 attributes are meaningful measures for “evidence-based” guidelines, but they are not natural or proper quality assessment considerations for performance guidelines.

Following this theory that the four evidence evaluations are irrelevant to guidelines in the group I designated as CG3, I removed them from the assessment rubric for this group, replacing them with two, more relevant, validity attributes: expert reference, and appropriateness criteria.

I further tailored the instrument to assess usability of this particular group of documents. In contrast to guidelines that provide “if-then” recommendations, performance guidelines offer multi-step directions, so in addition to the two new validity attributes I added these two items to assess usability: clearly presented steps, and effective document design.

To test my category-adjusted assessment approach, I evaluated six ultrasound guidelines. After assigning each guideline to a category, I rated each attribute and noted relevant text features. This detailed analysis revealed areas of document weakness that undermine the validity and usability of the guidelines, but it also illuminated writing approaches that strengthen them and add value. Regarding the category adjustments, I think they improve the rubric, reducing the instances of non-applicability. One exception is “Management Options.” I retained that attribute as universal, but for most of the guidelines in the group I analyzed, it was not truly relevant, and the scores I assigned and the examples I provided for most were arguably inflated. Another item that needs reconsideration is the attribute of “Logical presentation.” The term is too vague and lent very little value to the assessments.

My larger conclusion is that, while I believe tailoring assessment based on overall guideline purpose improves the tool’s sensitivity, its application needs to be refined. CPGs are complex documents, and most or all of the CPGs I reviewed contain sections that fit more than one of the guideline categories I identified. Based on my experience with this study, I conclude that for a truly legitimate analysis, a guideline’s various sections should be analyzed individually. Global features, like describing an updating procedure or providing an overall guideline objective, should be addressed in all guidelines as simple, clear statements, but other attributes, particularly those that relate to evidence, are more qualitative and complex. In contrast to items like “updating procedure” which is a stand-alone statement, the qualitative attributes themselves

relate to statements. As an example, it seems likely that in the near future performance guidelines will need to build appropriateness justifications. While the performance portion of the guideline – its body - may not need a strict evidence base, perhaps appropriateness justifications will require it and should be judged according to the four evidence-related criteria.

### **Is it appropriate to develop guidelines for these applications?**

One opinion is that guidelines should be written only for topics with a certain level of available evidence. Certainly unfounded recommendations should not be advanced, but this paper has included examples of strong recommendations made in the absence of high quality randomized controlled trials. Flu vaccination is one example of a treatment for which waiting to perform RCTs before writing guidelines would oppose patients' best interests. In the case of ultrasound performance guidelines, the justification for proceeding without "evidence" is a little different, but I believe valid. I contend that as practices emerge and early adopters develop skill sets and knowledge worthy of being shared by others, the expertise of those individuals is sufficient support for a guideline. In all cases, however, the critical and incontrovertible caveat is that the guideline's foundation needs to be fully and accurately conveyed by the guideline drafters.

### **If so, how can these guidelines best be constructed for validity and usability?**

The Guidelines International Network of CPG developers, and others, recognize AGREE II not only as a reliable quality assessment tool, but also as a useful guideline development tool. Following that precedent, I developed a template for devising guidelines for emerging ultrasound applications. The template is based on my revision of the AGREE II instrument, with special



consideration for the validity and usability challenges inherent in developing guidelines for emerging ultrasound applications.

*Clinical Practice Guideline (CPG)  
DEVELOPMENT TEMPLATE*

*The quality of clinical practice guidelines depends on their validity and usability. Following the principles described in this template will help guideline drafters promote those attributes.*

*Standard Components*

Eight standard text components should appear in every CPG. For each of these components, guideline drafters should meet the following reporting standards:

- 1) Objectives:
  - Clearly state guideline objectives and rationale.
- 2) Health question(s):
  - Explicitly word and prominently display the health question(s) to be addressed.
- 3) Target population:
  - Describe the target population using humanizing language.
- 4) Target users:
  - Specify all potential user groups.
- 5) Schedule and procedure for updating:
  - Report the publisher's revision schedule and procedure.
- 6) Implementation:
  - Provide implementation tools, or reference any related tools not described in the document.
- 7) Monitoring/audit criteria:
  - Supply monitoring and audit criteria, and identify areas for future research.
- 8) Conflicts of interest:
  - Disclose the presence or absence of competing interests for all members of the CPG development group.

*Specialized Validity Attributes*

*Two validity attributes should appear in most or all CPGs, but their features will vary according to the purpose of the guideline.*

1) Recommendations:

- Describe the guideline collaboration process, disclosing areas of disagreement, the method used to resolve differences, and the areas where differences remained unresolved.

*For recommendations based on evidence from the literature:*

- Outline the evidence selection criteria.
- Detail strengths and limitations of the evidence.
- Provide unambiguous links between evidence and recommendations.

*For recommendations based on expert opinion or consensus:*

- Identify the principal drafter and other content experts.
- List content experts' credentials: professional affiliations, experience, and publication credits.

2) Appropriateness:

- For evidence-based option stratification, appropriateness is intrinsically established.
- For all others, provide evidence of cost-benefit outcomes for applying the <recommended procedure> to answer the <health question(s)> for the <target population> by the <target users>.

*Specialized Usability Attributes*

*Guideline document usability features should suit the guideline's purpose.*

When documenting evidence-supported recommendations:

- Highlight and prominently display the health questions and recommendations.
- Provide intuitively arranged explicit links between clinical questions, relevant evidence, and recommended practices.

When describing procedures in a quick reference guide:

- List concise, directive steps.
- Begin each step on a new line.
- Group steps under exam task headings.
- Provide illustrations as needed.

When describing procedures in an extended version:

- Add technical tips and explanations of the purpose or goal of each step as indicated.

Figure 3: CPG Development Template

**APPENDIX A:  
ACCREDITATION**

Alfred Abuhamad, Beryl Benaceraf, Paula Woletz, and Bonnie Burke investigated the impact of an accreditation program on guidelines implementation, randomly selecting study subjects from among obstetric/gynecology practices accredited through the American Institute of Ultrasound in Medicine (AIUM) (Abuhamad 2004).

As part of the AIUM accreditation application, practices submit ultrasound case studies completed by their practitioners. Submitted obstetric and gynecologic studies are reviewed and scored for completeness according to criteria defined by AIUM Practice Guidelines for the Performance of Obstetric Ultrasound Examinations and for the Performance of the Ultrasound Examination of the Female Pelvis. To maintain accreditation, every three years practices must repeat the study submission process for reaccreditation.

To evaluate whether participating in the AIUM accreditation program affects practice, for each practice included in this research, scores on reaccreditation case study submissions were compared to scores assigned on studies submitted at the time of initial accreditation.

Results of this investigation confirmed that participants in the program improved scores in case studies submitted for reaccreditation. The authors propose that improved scores imply greater compliance with practice guidelines, “which should translate into enhancement of the quality of ultrasound practice.” These findings lend support to my premise that emerging users will benefit their patients by developing practice guidelines for their new applications (Abuhamad 2004).

**APPENDIX B:  
CPG QUALITY ASSESSMENT AND CONSTRUCTION TOOLKIT**

*The quality of clinical practice guidelines depends on their validity and usability. Following the principles described in this template will help guideline drafters promote those attributes.*

<b>Clinical Practice Guideline (CPG) ASSESSMENT TEMPLATE</b>						
Guideline Title:						
Guideline Developer:						
Guideline Category(s): GC1 _____ GC2 _____ GC3 _____						
<b>Instructions:</b>						
Rate Items 1 – 12 for ALL Guidelines. 1 = Strongly Effective 2 = Moderately Effective 3 = Moderately Ineffective 4 = Strongly Ineffective, or Missing GC1 and GC2 Guidelines: Rate Items 13 – 16 * GC3 Guidelines : Rate Items 17 – 20 **						
#	ATTRIBUTE	RATING				COMMENTS
		1	2	3	4	
1	Guideline objectives					
2	Health question(s)					
3	Target population					
4	Target users					
5	Updating procedure					
6	Recommendations					
7	Management options					
8	Key recommendations					
9	Implementation tools					
10	Monitoring, audit criteria, and future research					
11	Competing interests of CPG development group					
12	Logical presentation					
<b>*Rate Items 13 – 16 for CG1, CG2 guidelines ONLY!</b>						
13	Evidence selection criteria					
14	Evidence strengths, limitations					
15	Method for formulating recommendations					
16	Link between evidence & recommendations					
<b>**Rate Items 17 – 20 for CG3 guidelines ONLY!</b>						
17	Appropriateness criteria					
18	Clearly defined steps					
19	Effective document design					
20	Expert reference					
<b>TOTAL SCORES:</b>						= _____

## Clinical Practice Guideline (CPG) CONSTRUCTION TEMPLATE

### Standard Components

Eight standard text components should appear in every CPG. Following these reporting standards will enhance guideline quality:

- 1) Objectives:
  - Clearly state guideline objectives and rationale.
- 2) Health question(s):
  - Explicitly word and prominently display the health question(s) to be addressed.
- 3) Target population:
  - Describe the target population using humanizing language.
- 4) Target users:
  - Specify all potential user groups.
- 5) Schedule and procedure for updating:
  - Report the publisher's revision schedule and procedure.
- 6) Implementation:
  - Provide implementation tools, or reference any related tools not described in the document.
- 7) Monitoring/audit criteria:
  - Supply monitoring and audit criteria, and identify areas for future research.
- 8) Conflicts of interest:
  - Disclose the presence or absence of competing interests for all members of the CPG development group.

### Specialized Validity Attributes

*Two validity attributes should appear in most or all CPGs, but their features will vary according to the purpose of the guideline.*

- 1) Recommendations:
  - Describe the guideline collaboration process, disclosing areas of disagreement,



the method used to resolve differences, and the areas where differences remained unresolved.

*For recommendations based on evidence from the literature:*

- Outline the evidence selection criteria.
- Detail strengths and limitations of the evidence.
- Provide unambiguous links between evidence and recommendations.

*For recommendations based on expert opinion or consensus:*

- Identify the principal drafter and other content experts.
- List content experts' credentials: professional affiliations, experience, and publication credits.

2) Appropriateness:

- For evidence-based option stratification, appropriateness is intrinsically established.

For all others, provide evidence of cost-benefit outcomes for applying the <recommended procedure> to answer the <health question(s)> for the <target population> by the <target users>.

### **Specialized Usability Attributes**

*Guideline document usability features should suit the guideline's purpose.*

When documenting evidence-supported recommendations:

- Highlight and prominently display the health questions and recommendations.
- Provide intuitively arranged explicit links between clinical questions, relevant evidence, and recommended practices.

When describing procedures in a quick reference guide:

- List concise, directive steps.
- Begin each step on a new line.
- Group steps under exam task headings.
- Provide illustrations as needed.

When describing procedures in an extended version:

- Add technical tips and explanations of the purpose or goal of each step as indicated.

## LIST OF REFERENCES

- Abuhamad, Alfred Z., Beryl R. Benacerraf, Paula Woletz, and Bonnie L. Burke. "The Accreditation of Ultrasound Practices." *Journal of Ultrasound in Medicine* 23 (2004): 1023-1029. Print.
- Agency for Healthcare Research and Quality. Glossary of Terms. 9 September 2010. <http://www.effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms/>.
- Antman, Elliott M., and Raymond J. Gibbons. "Clinical Practice Guidelines and Scientific Evidence." *JAMA*. 2009;302(2):143-144.
- Arrive, Lionel. "Declining Number of Publications by American Authors in Radiology." *U6666*, 246 (2008): 330-331.
- Audet, Anne-Marie, Sheldon Greenfield, and Marilyn Field. "Medical Practice Guidelines: Current Activities and Future Directions." *Annals of Internal Medicine*. 1990;113;709-714. MEDLINE. Web. 7 May 2009.
- Baker, Laurence C., Elliott S. Fisher, and John E. Wennberg. "Variations In Hospital Resource Use For Medicare and Privately Insured Populations In California." *Health Affairs* 27, no. 2 (2008): w123-w134 (published online 12 February 2008; 10.1377/hlthaff.27.2.w123)
- Battista, Renaldo N., and Matthew J. Hodge. "Clinical Practice Guidelines: Between Science and Art." *Can Med Assoc J*. 1993; 148.3.385-389.
- Battista, Renaldo, Matthew J. Hodge, and Paolo Vineis. "Medicine, Practice and Guidelines: The Uneasy Juncture of Science and Art." *Journal of Clinical Epidemiology* 48.7 (195): 875-880.
- Burke, Kenneth. *Language as Symbolic Action*. **YEAR**. *The Rhetorical Tradition*. Ed. Patricia Bizzell and Bruce Herzberg. Boston: Bedford/St. Martins, 2001. **PAGES**. Print.
- Carter, Anne. "Clinical Practice Guidelines." Editorial. *Can Med Assoc J* 1992;147 (11) p. 1649-1650. WEB 7 May 2009.
- Charlton, BG. "Management of science." *Lancet* 1993; 342: 99-101.
- Chew, Felix S. "The Scientific Literature in Diagnostic Radiology for American Readers: A Survey and Analysis of Journals, Papers, and Authors." *American Journal of Radiology*. 147:1055-1062, November 1986. ny4 p 1055

“Clinical guidelines: Some considerations.” Editorial. *European Journal of Cardiovascular Nursing* 7 (2008) 91-93. MEDLINE. Web. 7 May 2009.

Dartmouth Institute for Health Policy and Clinical Practice. *Dartmouth Atlas of Health Care*. The Trustees of Dartmouth College. 2010. Web. 15 November 2010.

Douglass, Frederick. *Narrative of the Life of Frederick Douglass, an American Slave, Written by Himself*. 1845. *The Rhetorical Tradition*. Ed. Patricia Bizzell and Bruce Herzberg. Boston: Bedford/St. Martins, 2001. 1070-1075. Print.

Eddy, D.M. “Clinical Policies and the Quality of Clinical Practice,” *New England Journal of Medicine* 307, no. 6 (1982): 343–347

---. "Practice Policies: Where Do They Come From?" *Journal of the American Medical Association* 263, no. 9 (1990): 1265, 1269, 1272, 1275.

Field MJ, Lohr KN (eds): *Clinical Practice Guidelines: Directions for a New Program*, Committee to Advise the Public Health Service on Clinical Practice Guidelines, Institute of Medicine, US Dept of Health and Human Services, Natl Acad Pr, Washington, 1990: 38

*Guidelines for Clinical Practice: from Development to Use*, Committee on Clinical Practice Guidelines, Division of Health Care Services, Institute of Medicine, Natl Acad Pr, Washington, 1992: 37-38

Garland, L.H. *Annual Review of Medicine*. Vol 2: 273-290. Vol pub date 2/1951.

Goolsby, Mary Jo. “Evaluating and Applying Clinical Practice Guidelines.” *Journal of the American Academy of Nurse Practitioners* 13.1 (2001): 3-6. Print.

Grilli, Roberto, Nicola Magrini, Angelo Penna, Gorgio Mura, and Alessandro Liberati. “Practice Guidelines Developed by specialty societies: the need for a critical appraisal.” *The Lancet*. Vol 355. January 8, 2000. p. 103-106. Print.

Grimshaw, Jeremy, and Ian Russell. “Achieving health gain through clinical guidelines. I: Developing scientifically valid guidelines.” *Quality in Health Care* 1993 2: 243-248. group.bmj.com. Web. 5 June 2010.

---. “Achieving health gain through clinical guidelines. II: Ensuring guidelines change medical practice.” *Qual. Saf. Health Care* 1994;3;45-52. qshc.bmj.com. Web. 7 May 2009.

Grimshaw, Jeremy, N. Freemantle, S. Wallace, I. Russell, B. Hurwitz, I. Watt, A. Long, and T. Sheldon. “Developing and implementing clinical practice guidelines.” *Quality in Health Care* 4 (1995): 55-64.

- Guyatt, G.H. "Evidence-based Medicine," *ACP Journal Club* 114, no. 2 (1991): A-16. Evidence-based Medicine Working Group, "Users' Guides to the Medical Literature: I. How to Get Started," *Journal of the American Medical Association* 270, no. 17 (1993): 2093–2095.
- Guyatt, GH, D. L. Sackett, J. C. Sinclair, et al. Users' guides to the medical literature. IX: a method for grading health care recommendations. *JAMA* 1995; **274**: 1800–04.
- Institute of Medicine. *Medicare: a strategy for quality assurance*. Ed. Kathleen N. Kathleen. Washington, DC: National Academy Press, 1990. Print.
- Institute of Medicine. *Guidelines for clinical practice: from development to use*. Washington, DC: National Academic Press, 1992.
- McAlister, Finlay A, Sean van Diepen, Rajdeep S. Padwal, Jeffreu A:.. Johnson, Sumit R. Majundar. "How Evidence-Based Are the Recommendations in Evidence-Based Guidelines?" *PLOS Medicine* vol 4 issue 8 2007. e250. Web 15 November 2010.
- Richards, Ivor A., and Charles K. Ogden. *The Meaning of Meaning*. **YEAR**. *The Rhetorical Tradition*. Ed. Patricia Bizzell and Bruce Herzberg. Boston: Bedford/St. Martins, 2001. **PAGES**. Print.
- Sackett, D. L., et al "Evidence-based Medicine: What it is and What it isn't. (editorial) *British Medical Journal* 312 no 7023 (1996):71-72.
- Schuster, Mark A., Elizabeth A. McGlynn, and Robert H. Brook. How Good is the Quality of Health Care in the United States? *Milbank Quarterly*. Vol 76, Issue 4, pages 517-563/ December 1998.
- Shaneyfelt and Centor, "Reassessment of Clinical Practice Guidelines: Go Gently Into That Good Night," *JAMA*.2009; 301: 868-869
- Shekelle, Paul G, Steven H. Woolf, Martin Eccles, and Jeremy Grimshaw. "Clinical guidelines: Developing guidelines." *BMJ* 318 (1999): 593-596.
- The AGREE Collaboration. Appraisal for Guidelines for Research and Evaluation (AGREE) Instrument. [www.agreecollaboration.org](http://www.agreecollaboration.org). Web. 15 November 2010.
- United States. Centers for Disease Control and Prevention. *Updating the Guideline Methodology of the Healthcare Infection Control Practices Advisory Committee (HICPAC)*. Atlanta, GA. Web. 15 November 2010.
- "Validity." Def. 1. *Dorland's Illustrated Medical Dictionary*. 29<sup>th</sup> ed. 2000. Print.

Woolhandler, Steffie, Terry Campbell, and David U. Himmelstein. "Cost of Health Care Administration in the United States and Canada." *The New England Journal of Medicine* 349;8 (2003); 768-775.

Workshop on Clinical Practice Guidelines: summary of proceedings. Department of Health Care and Promotion, Canadian Medical Association CAN MED ASSOC J 1993; 148 (9) MAY 1, 1993. 1715-1719.