Lack of Uniformity in Levels of Evidence and Recommendation Grades in Surgical Oncology Guidelines

Haejin In · Caprice C. Greenberg

Published online: 28 April 2012
© Société Internationale de Chirurgie 2012

Practice guidelines exist to bridge the gap between research and clinical practice. “Ensuring Quality Cancer Care” published in 1999 by the Institute of Medicine’s National Cancer Policy Board [1] promoted practice guidelines as a means to improve the quality of care and brought about a flurry of activity with various societies producing practice guidelines for oncology. These various organizations used a myriad of methods to grade the level of evidence (LOE) and strength of recommendations [2], resulting in both nonuniform expression of and inconsistency in conclusions. As guidelines are used increasingly to create performance indicators to measure quality of care [3], the stakes are higher, making variability in guideline recommendations problematic.

Eight cancer-specific guidelines generated by organizations in the United States and applicable to surgical oncology were selected to examine the rating schemes and processes used to assign LOE and grade of recommendation. Table 1 describes the guideline organizations

considered. For brevity, the full name is spelled out in Table 1 and the text refers to organizations by acronym.

A guideline recommendation is developed generally through two distinct steps and expressed through organization-specific rating schemes. First, the LOE is evaluated. This is done through an objective determination of a level according to the study design type used in the supporting evidence and a subjective interpretation of the evidence by panel members about the scientific rigor of the studies. Organizations were inconsistent about which inputs to use in determining LOE; whereas some organizations used only type of study design or the subjective judgment of the panel members, some organizations used a combination of both. Organizations that used type of study design as its criteria for level determination readily documented a LOE, whereas organizations that predominantly used subjective judgment documented LOE less frequently.

Second, a recommendation grade (RG) is determined. The grade expresses how strongly a guideline organization wishes to advocate a recommendation. A guideline panel will generally assign a grade using a combination of the LOE and the judgment and opinion of the expert panel about the practice and policy implications of a recommendation. While LOE was the prevailing consideration in many guidelines (AACE (2004), ASCO (2001), ASCRS (2005)), other guideline organizations favored the use of subjective judgment regarding the quality of the evidence (ATA, CAP, NCCN) to determine RG.

LOE or RG was not used by all guideline organizations. Although three of the eight guidelines used both, two ranked only LOE, three ranked only RG, and two of the guidelines ranked neither LOE nor RG (Table 1). The annotation used by guideline organizations also was variable. Whereas numeric annotation for LOE and alphabetic annotation for RG was most common, this was not always

H. In · C. C. Greenberg
Department of Surgery, Center for Surgery & Public Health,
Brigham and Women’s Hospital, Boston, MA, USA

H. In
Institute for Technology Assessment, Massachusetts General
Hospital, Boston, MA, USA

H. In (✉)
Department of Surgery, University of Chicago Medical Center,
Chicago, IL, USA
e-mail: Haejin.In@uchospitals.edu

C. C. Greenberg
Department of Medical Oncology, Center for Outcomes &
Policy Research, Dana-Farber Cancer Institute, Boston, MA,
USA

Table 1 Heterogeneity of annotation schemes used to express LOE and RG

	LOE	RG
American Association of Clinical Endocrinologists (AACE) Guidelines (2004)	I–4	A–D
American Society of Clinical Oncology (ASCO) Guideline Recommendations (2001)	I–V	A–D
American Society of Clinical Oncology (ASCO) Guideline Recommendations (2004)	–	–
American Society of Colon and Rectal Surgeons (ASCRS), Practice Parameters (2004)	I–III, A–C	–
American Society of Colon and Rectal Surgeons (ASCRS), Practice Parameters (2005)	I–V	A–D
American Thyroid Association (ATA) Guidelines	–	A–I
College of American Pathologists (CAP) Consensus Statement 1999	–	Category I–IV
National Comprehensive Cancer Network (NCCN) Practice Guidelines in Oncology	–	Category 1–3
National Cancer Institute (NCI) Treatment Guidelines	Ii–3iii	–
Society for Surgery of the Alimentary Tract (SSAT) Patient Care Guidelines	–	–

the case. Additionally, LOE and RG were sectioned in various ways, ranging from three tiers (I to III) to six tiers (Ii to 3iii). Although a unidirectional scale was generally used to rank levels of evidence, either unidirectional or bidirectional scale could be used to express grades of recommendation. Meta-analysis and systemic reviews of randomized control trials (RCT) were considered by all organizations to provide the highest LOE, whereas expert opinion was considered to provide the lowest LOE. However, the rating for study designs between the two extremes varied significantly by organization (Table 2).

Almost all guideline organizations used rigorous and explicit methodologies to evaluate the evidence and develop guideline recommendations. However, the lack of consistency in the usage of LOE and RG, the nonuniformity of inputs used to determine LOE and RG, and the diversity of rating systems used to express LOE and RG raise two concerns. First, a lack of uniform methodology to document expressly the quality of evidence used to create a recommendation undermines the utility of guidelines as a tool to narrow the evidence to practice gap. Studies that examine physician behavior show that physicians will change practice to accept recommendations in response to good evidence [4, 5]. However, guidelines often offer inconsistent [6] or even

conflicting advice [7], complicating integration into practice. Lack of clear documentation exacerbates this concern and throws further doubt on its reliability.

A second concern regards the increased use of guideline recommendations as quality and performance measures to evaluate and rank institutions and physicians. The lack of a set criterion to determine and express LOE and RG make it difficult to justify using a particular organization's guideline recommendation. Given the ongoing call for quality measures by the public and payers, improving the standardization across guidelines takes on an increased urgency.

Prior efforts have mostly focused on the evaluation of guidelines ("guidelines for guidelines") and less on standardization. There are at least 24 different instruments available for assessing the quality of clinical practice guidelines [8, 9], including the Conference on Guideline Standardization (COGS) and the Appraisal of Guidelines, Research and Evaluation (AGREE) collaboration. Efforts promoting standardization have been scarce, but a universal approach to grading the quality of evidence and strength of recommendations developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group has gained momentum [10]. The AACE (version 2010) and various other organizations,

Table 2 Varying levels of evidence assigned by study design

	Meta-analysis, systemic reviews of RCTs	Multiple RCTs	Single RCT	Low quality RCTs (high likelihood of bias)	Multiple cohort studies	Individual cohort studies	Multiple case– control studies	Individual case– control studies	Time series	Case series/ descriptive studies	Expert opinion
AACE (2004)	I	I	2	3	2	2/3	3	3		3	4
ASCO (2001)	I	I	II	II	III	III	III	III	III	IV	V
ASCRS (2004)	I (A/B)	I (A/B)	I (A/B)	II (A/B/C)	II (A/B/C)	II (A/B/C)	II (A/B/C)	II (A/B/C)	II (A/B/C)	III (C)	III (C)
ASCRS (2005)	I	I	II	II	III	III	III	III	III	IV	V
NCI	1	1	1	2	3	3	3	3	3	3	–

AACE American Association of Clinical Endocrinologists, ASCO American Society of Clinical Oncology, ASCRS American Society of Colon and Rectal Surgeons, NCI National Cancer Institute

such as American College of Chest Physicians and National Institute for Clinical Excellence, recently have incorporated the GRADE methodology.

Consumers of medical information increasingly are flooded by information from multiple sources and are forced to decide what they should and should not adopt into practice and policy. Guidelines initially created to aid in this process may instead complicate matters as they increase in number and offer inconsistent conclusions. A systemic and uniform method to develop and present LOE and RG that allow comparisons across different guidelines would (1) enhance usability by providing credence to recommendations, (2) promote the use of guidelines as recommendations to inform practice, and (3) enhance physician buy-in for the use of quality measures derived from practice guidelines. Such an approach would aid efforts to align policy and protocols with clinical practice as well as administrative incentives with good medicine.

References

1. Moulton G (1999) IOM report on quality of cancer care highlights need for research, data expansion. *Institute of Medicine. J Natl Cancer Inst* 91(9):761–762
2. Schunemann HJ, Fretheim A, Oxman AD (2006) Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Res Policy Syst* 4:21
3. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A et al (2003) The quality of health care delivered to adults in the United States. *N Engl J Med* 348(26):2635–2645
4. Stafford RS, Furberg CD, Finkelstein SN, Cockburn IM, Alehegn T, Ma J (2004) Impact of clinical trial results on national trends in alpha-blocker prescribing, 1996–2002. *JAMA* 291(1):54–62
5. Simunovic M, Baxter NN (2009) Knowledge translation research: a review and new concepts from a surgical case study. *Surgery* 145(6):639–644
6. Swales JD (1993) Guidelines on guidelines. *J Hypertens* 11(9): 899–903
7. Thomson R, McElroy H, Sudlow M (1998) Guidelines on anti-coagulant treatment in atrial fibrillation in Great Britain: variation in content and implications for treatment. *BMJ* 316(7130): 509–513
8. Oxman AD, Schunemann HJ, Fretheim A (2006) Improving the use of research evidence in guideline development: 16. Evaluation. *Health Res Policy Syst* 4:28
9. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D (2005) A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care* 17(3):235–242
10. Calonge N (2010) New promise for uniform evidence-based guideline development: The GRADE Approach. <http://www.guideline.gov/expert/commentary.aspx?file=GRADE.inc>. Accessed 13 June 2010