SCIENTIFIC ARTICLE

Quality Measures That Address the Upper Limb

Hand Surgery Quality Consortium*

Purpose Physicians, health care systems, and payers use quality measures to judge performance and monitor the outcomes of interventions. Practicing upper-limb surgeons desire quality measures that are important to patients and feasible to use, and for which it is fair to hold them accountable.

Methods Nine academic upper-limb surgeons completed a RAND/University of California–Los Angeles Delphi Appropriateness process to evaluate the importance, feasibility, and accountability of 134 quality measures identified from systematic review. Panelists rated measures on an ordinal scale between 1 (definitely not valid) and 9 (definitely valid) in 2 rounds (preliminary round and final round) with an intervening face-to-face discussion. Ratings from 1 to 3 were considered not valid, 4 to 6 were equivocal or uncertain, and 7 to 9 were valid. If no more than 2 of the 9 ratings were outside the 3-point range that included the median (1–3, 4–6, or 7–9), panelists were considered to be in agreement. If 3 or more ratings of a measure were within the 1 to 3 range whereas 3 or more ratings were in the 7 to 9 range, panelists were considered to be in disagreement.

Results There was agreement that 58 of the measures are important (43%), 74 are feasible (55%), and surgeons can be held accountable for 39 (29%). All 3 thresholds were met for 33 measures (25%). A total of 36 reached agreement for being unimportant (48%) and 57 were not suited for surgeon accountability (43%).

Conclusions A minority of upper-limb quality measures were rated as important for care, feasible to complete, and suitable for upper-limb surgeon accountability.

Clinical relevance Before health systems and payers implement quality measures, we recommend ensuring their importance and feasibility to safeguard against measures that may not improve care and might misappropriate attention and resources. *(J Hand Surg Am. 2016;■(■):■–■.*

Key words Hand surgery, performance measure, quality, quality measure, value.

Additional Material at jhandsurg.org

NEW HEALTH CARE LEGISLATION IN the United States aims to shift reimbursement from volume to value.[1,2] The assessment of value depends on measures of quality and cost. Quality measures are tools used to evaluate the degree to which appropriate care is provided.[3–7] They are used by physicians, health care systems, and payers to judge the performance of facilities and physicians, and to monitor the effects of interventions.[8–14] They are also used to assess "best practice" behaviors to allow for comparison of the quality of care delivered between health care providers and institutions.[3] Ideally they are

*Hand Surgery Quality Consortium members include: Robin N. Kamal, MD, David Ring, MD, PhD, Edward Akelman, MD, David S. Ruch, MD, Marc J. Richard, MD, Amy Ladd, MD, Chris Got, MD, Phil Blazar, MD, Jeff Yao, MD, and Sanjeev Kakar, MD.

Received for publication February 16, 2016; accepted in revised form July 18, 2016.

No benefits in any form have been received or will be received related directly or indirectly to the subject of this article.

Corresponding author: Robin N. Kamal, MD, Department of Orthopaedic Surgery, Stanford University, 450 Broadway Street, Pavilion A, Redwood City, CA 94063; e-mail: rnkamal@stanford.edu.

0363-5023/16/■ ■-0001$36.00/0
http://dx.doi.org/10.1016/j.jhsa.2016.07.107

evidence based, important for patient care, and feasible for both the physician and health system.[15,16] For example, the percentage of patients who received appropriate antibiotic prophylaxis of infection before hip or knee arthroplasty is a good quality measure because it has supporting evidence, is important to patients because it lowers the risk of infection, and is feasible to measure, quantify, and track.[17,18]

Quality measures most often assess the degree to which management follows accepted practices of care (process measures) or improves health status (outcome measures).[19] However, some widely implemented quality measures do not improve patient care or outcomes.[20–22] The success of efforts to improve the value of health care will depend to some degree on agreement among stakeholders that the selected quality measures are appropriate and worthwhile.[3,6,23] If this agreement does not exist, there is a risk of creating health care systems that are optimized to perform well on metrics but do not actually improve the health of patients. When health care payments are based on adhering to quality measures, the question of who should be held accountable for ensuring compliance with these measures also becomes an issue. For example, should the physician be penalized when a patient falls in the hospital after surgery, or should the hospital be penalized? This study assessed the importance, feasibility, and accountability of current upper-limb quality measures using a validated method.

MATERIALS AND METHODS

Identification of quality measures

We previously completed a systematic review and found 134 quality measures addressing the upper limb[24] (Appendix A; available on the *Journal's* Web site at www.jhandsurg.org). Briefly, measures were found following Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines[25] using MEDLINE/PubMed, EMBASE, Google Scholar, American Academy of Orthopaedic Surgeons Clinical Practice Guidelines, the National Quality Forum, the Agency for Healthcare Research and Quality, and the Centers for Medicare and Medicaid Services Physician Quality Reporting System (2015). One panel member (D.R.) was involved in developing the American Academy of Orthopaedic Surgeons Clinical Practice Guidelines on Distal Radius Fractures that was used for this study.

Quality measure evaluation

We completed a modified RAND/University of California–Los Angeles (UCLA) Delphi Appropriateness process of 134 quality measures using a 9-person panel of academic hand surgeons (Hand Surgery Quality Consortium) to evaluate the importance, feasibility, and accountability of current quality measures that address the upper limb. The Hand Surgery Quality Consortium is composed of surgeons and quality measure development experts with a mission of studying quality in upper-limb surgery. We evaluated quality measures using the Agency for Healthcare Research and Quality methodology for assessment of candidate indicators.[26] All members involved in the panel are board-certified orthopedic surgeons with additional training in hand and upper-limb surgery and individual practices that include adult, pediatric, elbow, and shoulder problems; are in rural and urban practices; and are located geographically in the west, midwest, south, and east. We followed the RAND/UCLA Appropriateness methodology, which is a well-established consensus process that includes a synthesis of findings from the literature and iterative panel ratings, and that produces appropriateness criteria and quality measures that have face, construct, and predicative validity.[27–36] The methods used in this study followed the methods of prior studies, including measurement rating and categorization of data, analysis of results, and employment of 2 rounds with a face-to-face meeting between rounds.[37–41] Panelists were provided with the list of current quality measures (Appendix A; available on the *Journal's* Web site at www.jhandsurg.org) and their supporting literature, along with the definition of terms[16] (Table 1). They were asked independently to rate the validity of each quality measure based on 3 criteria[26]: (1) importance for clinical care, (2) feasibility of completing the measure in practice, and (3) whether hand or upper-limb surgeons should be held accountable for adhering to the measure.

RAND/UCLA Delphi scoring

The RAND/UCLA Delphi process includes 2 rounds of independent ratings of quality measures (preliminary and final), with a face-to-face group discussion between rounds (Fig. 1).[28,38–41] Panelists are asked to rate each measure on a scale of 1 to 9 in which 1 = definitely not valid, 5 = uncertain or equivocal validity, and 9 = definitely valid. Ratings from 1 to 3 are considered not valid, 4 to 6 indicates equivocal validity or uncertain, and 7 to 9 indicates validity. Panelists are informed of the scoring process and how the ordinal data will be categorized before voting. Scoring was completed according to the RAND/UCLA process. The first (preliminary) and second (final) rounds of ratings were used to calculate a median score of each quality measure.